Sampling in Forest Surveys 3

3.1 Introduction

Inductive reasoning based on assessing a part of a whole is very much part of our daily life. For example, when cooking a pot of spaghetti we infer the *al* dente quality of all noodles by checking just a few. Where "the whole" is diverse, complex, and extensive it might be risky to draw conclusions from only one or a few instances. Sampling is an example of inductive logic by which conclusions are inferred on the basis of a limited number of instances. A sample is a subset of a population, which itself is the entire set of elements for which estimates about specific characteristics are to be obtained.

In the context of forest resource assessments the collection of information by means of a complete enumeration is only feasible in exceptional situations. An alternative to complete enumeration is a sample survey, which serves as the basis for estimates or inference for the underlying population. The process of selecting a sample from a population is called sampling.

The first part of this chapter presents some basic terms and concepts, while the second part describes some sampling procedures important for forest resource assessments. For further reading the textbooks of Kish (1965), Cochran (1977), Sukhatme et al. (1984), de Vries (1986), Särndal et al. (1992), Thompson (1992), Schreuder et al. (1993), and Shiver and Borders (1996) are recommended.

3.2 Basic Concepts

3.2.1 Population, Samples, and Estimators

A population comprises all elements from which the sample is to be taken. It may be defined very simply, for instance, the trees of a forest stand or the participants of a workshop. The definition of the population must be unique and

allow an operational and comprehensible decision on whether a questionable element belongs to the population or not.

The population from which the sample is taken is termed the sampled population and must match the target population for which estimates are desired. Only then can representative conclusions for a target population be drawn. The population can be either infinite or finite depending on the definition. A population defined as the forest in a given region may comprise an infinite number of spatial locations but only a finite number of trees. We also need a temporal definition of a population. Few populations remain constant over time; most undergo changes due to birth, mortality, emigration and immigration processes. In the example before, the number of trees in a forest is likely to change over time. When a population is finite and countable it is customary to denote the population size by N, where N is a positive integer. For example, for a population composed of a single forest stand with 200 trees N=200.

A sample consists of a number of sampling units (or simply units) selected from the population by some design. The population is also uniquely defined in terms of these units as the union of all possible samples of such units. Sample units can be either unique discrete nonoverlapping units or arbitrarily sized units located at random in the population (Williams and Eriksson 2002). In the former case we view the population as a finite set of unique units that completely tesselate the population. The tesselated paradigm ensures that every population element can be clearly allocated to one unique unit. Examples of sample units are single trees, sample plots, or districts. In the second case we view the population as composed of an infinite set of possible locations for our sampling units. A sample location provides attribute information that is representative of the sampled locations.

For finite countable populations the N individual units of a population are identifiable, if they can be uniquely labeled from 1 to N and the label of each unit is known (Schreuder et al. 1993). While it may be relatively easy to identify and label N trees in a forest stand, the issue of identifiability quickly becomes an insurmountable logistic obstacle when the population of the trees in a large forested area. In extensive forest surveys the construction of an exhaustive list of sampling units, called the *sampling frame*, is often one of the major practical problems (Särndal et al. 1992). Without a complete sampling frame one must adopt the point paradigm for a definition of the population. Populations defined by the point paradigm are often less than intuitively clear.

A forest inventory sampling frame is often assembled from a unique description of what qualifies as forest (i.e., the forest area definition). Forest area definitions utilize quantitative criteria such as crown density, minimum patch size, or minimum patch width to facilitate a forest/nonforest decision in order to construct the sampling frame. Care must be given to the definitions to ensure that the qualifying population is indeed the population of interest. Even

minor changes in a definition may lead to substantial and often surprising changes to the qualifying population. The following example illustrates this phenomenon. A survey of two treed areas with 25 sample plot locations in each area (circles) arranged in a regular grid found that 13 sample locations were covered by a tree canopy in the first area while only four were covered in the second area (Fig. 3.1). Accordingly the sample-based estimates of the crown density in each area are 13/25 (52%) and 4/25 (16%), respectively. A forest qualifying threshold of 10% would result in a classification of both areas as forest. In contrast, only the first area would qualify as forest if the defining threshold was raised to 30%.

Each sample unit and population element possess a series of attributes of interest. The attribute may be intrinsic or derived. The cellulose content would be an example of an intrinsic property. A market value, on the other hand, is an example of a derived attribute – an attribute that can only be obtained via another attribute or several other attributes. Natural resource attribute values often exhibit a considerable variation between units (elements). An attribute may or may not be measurable or quantifiable. Instead we may define countable or measurable variables linked to the attribute of interest. Knowledge of and information on these variables are used for inference about the attributes of interest. For example, the attribute of interest may simply be the "trees" in a forest with the element attribute being "tree." To characterize this attribute beyond a mere count of trees we may choose to measure variables such as tree height and stem diameter at breast height, identify the tree species, and assess crown form. The number of variables to include will depend on what is needed to be known about the attribute of interest.

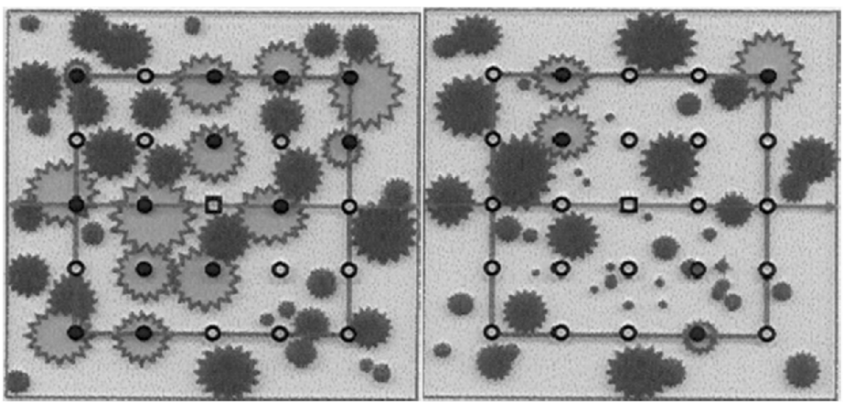


Fig.3.1. Two treed areas with 25 sample locations on a regular grid (*circles*). Sample locations under a tree canopy cover are indicated by *filled circles* (13 in the *leftmost area* and four in the *rightmost area*). (Courtesy of Markus Keller, WSL, Switzerland)

In order to characterize the attributes (variables) of a population certain parameters are employed. When the parameters relate to all units/elements in a population they are called population parameters. The aim of surveys is to estimate population parameters or the functions of one or more of them. The value of a parameter derived from a sample is called an estimate. The formula for calculating this estimate is called an estimator. Parameters include aggregates (e.g., total volume, total area) and averages (e.g., mean tree height) of values associated with each population element or unit. Ratios of pairs of population parameters (e.g., volume per hectare as the ratio of total volume and total area), counts (e.g., number of trees), and proportions (e.g., proportion of forest area with a specific attribute) are further examples of population parameters.

Estimates of population parameters are obtained via estimators. The estimators treated in this book are either design-based or model-based estimators (Särndal et al. 1992; Gregoire 1998; Little 2004). The underlying principle behind a design-based estimator is that the population from which samples are taken is considered as a fixed entity. The random selection of units/elements to include in the sample is the only source of stochastic variation (sampling error). Model-based estimators are based on the assumption that the population of interest is generated by some process, a process that depends on a set of parameters to be estimated from the sample. The actual population to be surveyed is but one random realization from this process. We cannot observe the assumed process, but our sample allows us to estimate the parameters of the assumed model. Population estimates are obtained by combining the sample estimates with the model-based predictions for the nonsampled part of the population (Valliant et al. 2000). Thus, the issue of model bias looms large over these estimators and convincing support for the chosen model must come from previous surveys or from the sample data themselves.

Parameters of a population are designated by capital letters from the Latin and Greek alphabets. Lowercase letters are reserved for individual unit/element values.

3.2.2 Probability Sampling

The general principle of sampling (Fig. 3.2) is to select a subset of units (i.e., a sample) from a population, to measure this subset intensively, and to draw inference from the sample to the entire population.

There exist countless approaches to select a sample from a population. Intuitively it is obvious that the sample should represent the entire population. The term representative as used in everyday language suggests that the sample should be a tailgate miniature or a scaled-down replica of the population.

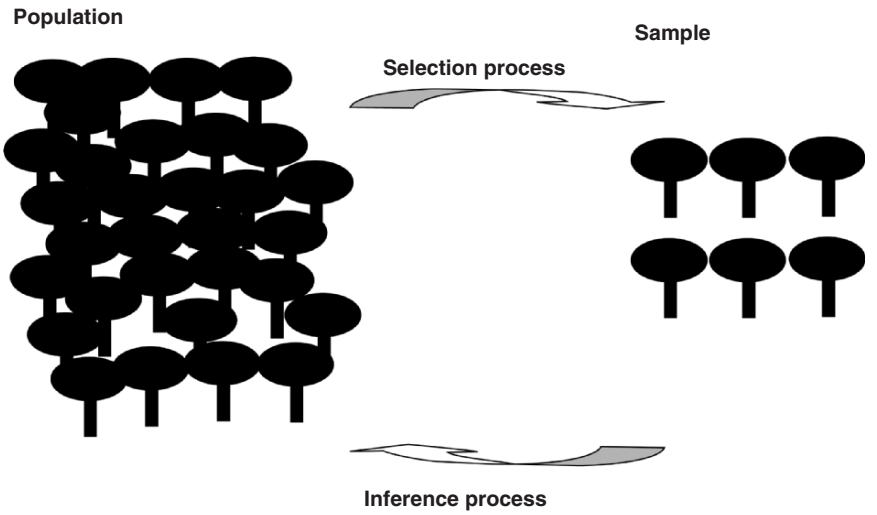


Fig. 3.2. The principle of sampling

Unless each unit in the population has an equal chance of being selected, this intuitive concept is inappropriate. Many widely used sampling methods assign varying selection probabilities to the individual units; the chance of being selected can be assigned with respect to a known attribute or quantitative measure of the units. A selection method complies with the conditions of probability sampling when a procedure is followed that ensures that each unit in the population has exactly the predetermined probability of being selected for the sample. The selection probabilities are used in the estimators of parameters of interest and in estimators of sampling variance (Thompson 1992). The choice of selection probabilities and estimators is called a sampling strategy (Särndal et al. 1992).

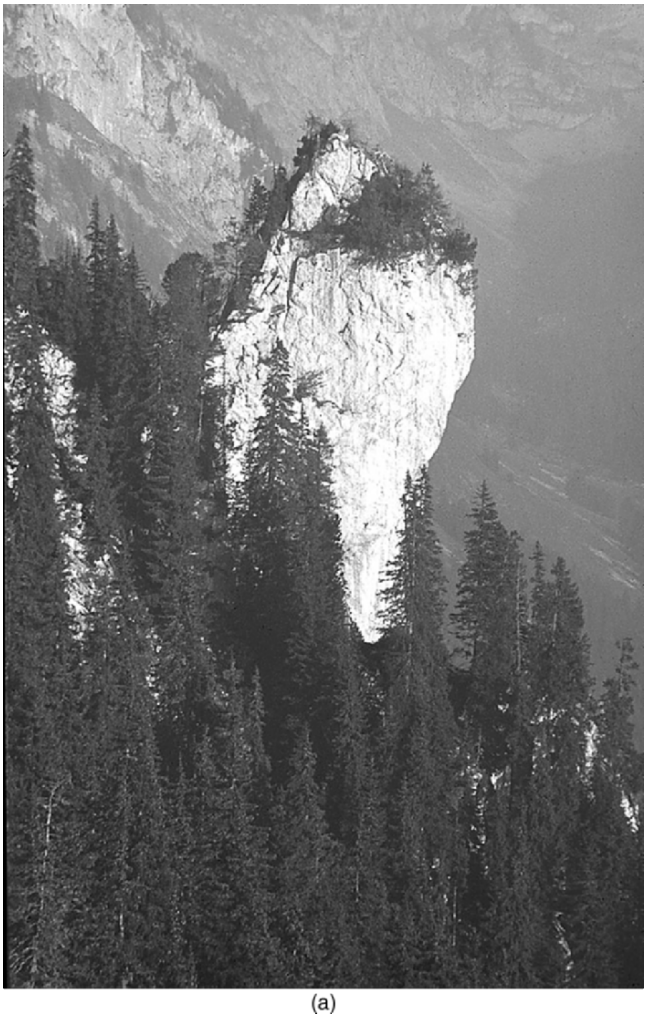
Given a specific population of N units the set of all possible distinct samples, s_1, s_2, ..., s_v, can be defined and the units making up each sample can be designated (two samples are distinct if their union minus their intersection is not empty). If n units out of N are to be selected without replacement (a unit can only be selected once) there are $\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$ possible distinct samples (Levy and Lemeshow 1991). Note, $n! = n(n-1)(n-2)\ldots(1)$ and 0!=1. For example, if a population contains 200 elements and we wish to take a sample of 25 elements, then the total number of possible samples is approximately 4.5×10^{31} (exact number is 45,217,131,606,152,448,808,778,187,283,008), quite an astronomical figure. For each possible sample, say s_i, a selection probability π_{si} can be specified. The sum of these selection probabilities over all possible

samples is 1. A selection probability tells us how frequent a particular sample s_i will be selected. We shall see later how we use these selection probabilities to derive unbiased estimators of population attributes and parameters. Further details on the use and significance of selection probabilities are in, for example, Brewer and Hanif (1983). The term probability sampling refers to sampling with a known selection probability of all sample units making up the population. In practice it is hardly possible to list all possible samples s_i and their associated selection probabilities π_{si}. For sample estimators based on a probability sample it is sufficient to know how to assign selection probabilities to the sampled units. We denote the actual sample by the symbol *s*, where *s* is one of the possible distinct samples. Estimates of population totals and averages are normally obtained by an *expansion* of individual sample attributes/variables to an estimate(s) of the population total (Levy and Lemeshow 1991). Let y_i be the attribute/variable obtained from sample *i* with selection probability π_i. The expanded estimate is y_i/π_i. Since the probability of obtaining this expanded estimate is π_i the expected value of the *i*th expanded value is simply y_i. Thus, expansion estimators of totals and averages are unbiased (if sampling was exhaustive the total, or average, would be equal to the true total, or average).

Probability sampling methods employ a thorough selection process that ensures that each unit in the population to be sampled has exactly its designated probability of being selected. In practice that means that any unit being selected as part of the sample has to be accepted, irrespective of any problems or difficulties in assessing it. A common problem in forest surveys is the accessibility of terrain. Figure 3.3 shows a forest patch that is inaccessible by a normally equipped field crew. Where these areas are not excluded from the sampling frame (i.e., inventory results refer to accessible forests only) they need to be surveyed when selected under a probabilistic sampling scheme. Often it is cheaper, quicker, or more comfortable to omit those units. This leads to in the problem that there is no longer control over the probability with which the units comprising the population are selected. Some units have little or no chance of being selected or are selected with uncontrolled or subjective probabilities. Such samples are called nonprobability samples.

3.2.3 Definitions and Notations

In sample surveys, data on one or more variables/attributes are collected for each selected unit of the population. Values reflecting a variable of a unit/element forming a finite population are defined by $y_i (i = 1, \ldots, N)$ and N is the number of population units/elements. Through sampling, a sample s composed



(a)

Fig. 3.3. Inaccessible forest areas in **a** Switzerland and **b** Germany

of n units is selected from the population units/elements. Variable/attribute values associated with a sampled unit are denoted by $y_{i|i \ni s}$, where $i \ni s$ means that the ith population value is sampled. To simplify notation we will drop $i \ni s$ when warranted. The ratio n/N specifies the proportion of units selected from the N population units and is termed the sampling fraction; the symbol f is often used to denote this fraction. In infinite populations the sample fraction is nil by definition.

(b)

Fig. 3.3. (Continued)

Sampled units are used to estimate parameters for the population. The four most important population parameters are:

1. The mean $\overline{Y}$ (e.g., the mean standing reserve in the inventory area)
2. The total Y (e.g., the total standing reserve in the inventory area)
3. The ratio R between two means or totals (e.g., volume per hectare)
4. The proportion P of units with a specific attribute (e.g., proportion of units with a given tree species).

The sample provides us with estimates of population parameters. Estimates are distinguished from their true population values by adding a caret above the associated symbol. The relationship between population values and sample estimators is given in Table 3.1 for the most common population parameters.

3.2.4 Properties of Estimators

Whereas the term "estimate" signifies the value of a parameter, an "estimator" denotes a rule according to which a parameter is derived from the sample data. Estimators based on sampling surveys must display certain qualities.

Table 3.1 Population and sample estimators of common parameters

Parameter	Population value	Sample estimator
Mean	$\bar{Y}$	$\hat{\bar{Y}} = \frac{1}{n}\sum_{i \ni S} y_i$
Total[a]	Y	$\hat{Y} = N \times \hat{\bar{Y}}$
Ratio[b]	R	$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\hat{\bar{Y}}}{\hat{\bar{X}}} = \frac{\sum_{i \ni s} y_i}{\sum_{i \ni s} x_i}$
Proportion	P	$\hat{P} = \frac{1}{n} \times \sum_{i \ni s} \delta_i$, *where* $\delta_i = \begin{cases} 1 \textit{ if the ith unit has the attribute} \\ 0 \textit{ otherwise} \end{cases}$

[a] finite populations only
[b] of means viz. totals

An estimator is called a "consistent estimator" if the larger the sample size n, the closer the estimate, say $\hat{Y}$, is to the true population parameter value Y. When the expected value of the estimator $E\left(\hat{Y}\right)$ equals the true parameter Y, the estimator is unbiased. Estimators not meeting this condition are termed "*biased estimators*." Bias is defined as the difference between the expected value of an estimator of a population parameter and the true value of this parameter. For the estimator $\hat{Y}$ the bias is given by $bias\left(\hat{Y}\right) = E\left(\hat{Y}\right) - Y$. An estimator that is unbiased for a given sample design (if correctly implemented) is said to be design-unbiased. Model-based estimators are said to be model-unbiased if the model is true and the expectations of model predictions equal the expectations of the population units for which predictions are made.

Unbiasedness is a desirable property of an estimator. Important is also the accuracy of an estimator. In repeated sampling of a single population using the same sampling design the estimates obtained from an estimator will vary between samples. Accuracy refers to the size of the deviations of the sample estimates from their true value (Cochran 1977). Normally, though, we would not know the true value. Different estimators of the same population parameter can have different accuracy. Normally, though, we do not know the true parameter value, which precludes a correct assessment of accuracy. We can, however, estimate the precision of an estimator. Precision is a measure of the deviations of individual sample estimates in repeat sampling from their mean (average). Precise estimators produce estimates that cluster tightly around their average. That means that we can have a high degree of confidence in the value of a sample-based estimate. If we were to repeat the sampling process we would likely obtain a result quite similar to the one we already have. Precision is commonly quantified as the inverse of the estimated variance of an estimate (Cochran 1977). To assess precision of an estimate we need an estimate of its sampling variance (var). Estimators of sampling variance have been developed

for all practically relevant sampling designs and population parameters including model-based predictions.

Bias and precision of estimators are both important attributes to consider in planning a survey. Estimators for various design alternatives (viz., model alternatives) may produce different amounts of bias and vary in precision. Actually, competing estimators often display a trade-off between bias and precision (Congdon 2001). The usual criterion for comparing two estimators is the mean square error (MSE). The MSE of an estimator, say $\hat{Y}$, is defined as

$$\mathrm{MSE}\left(\hat{Y}\right) = \mathrm{var}\left(\hat{Y}\right) + \mathrm{bias}\left(\hat{Y}\right)^2$$

Note, the true variance of a sample-based estimate and the true bias will never be known in practical applications. Instead we use available estimators of variance and approximations to the bias (Särndal et al. 1992). A survey analyst normally prefers an estimator with the lowest expected MSE.

Robust estimators are also desirable (Staudte and Sheather 1990). Robust estimators are less sensitive to a few outlying sample values and to violations of assumptions than are nonrobust estimators. The ideal estimator is unbiased, highly precise, and robust. It is a challenge for the survey analyst to optimize the sampling strategy, i.e., the choice of sample design and estimators.

3.3 Survey Design and Sampling Design

In planning a forest inventory, a range of methodological issues have to be considered. What data are to be collected? For which units of the population should they be collected? Which system of nomenclature (including measurement rules or definitions for each attribute to be assessed) is to be applied? How should data be captured and processed in order to derive the requested information? Additionally operational, organizational, and administrative issues have to be resolved. It is the objective of the *survey design* to settle these issues with respect to the available budget and the information needs. Ideally this process could be formulated as an optimization problem. What is the best design under a fixed set of resources and precision target? Examples of how this problem is resolved in various settings are found in, for example, Mandallaz 2001, Brus et al. 2002, and Arner et al. 2004). Historic material in the form of data from previous or related surveys provides a good source of prior information of what the intended sampling may produce.

A second set of methodological issues deals with the question of how to select the sample from the population (i.e., the sample selection) and how to derive suitable estimates from the data collected (i.e., the estimation procedures). Based on sampling theory, a variety of techniques have been developed for sample selection and estimation. It is the objective of the *sampling design* to select

the most appropriate sampling methods in light of a set of overarching objectives and constraints. The sampling design itself is part of the survey design.

Sampling designs can be divided into two main groups depending on whether data and information outside the variables of primary interest (auxiliary information) are used to shape the design and/or the estimators:

1. Sampling designs without auxiliary information
2. Sampling designs with auxiliary information

In sampling designs without auxiliary information, only the observations on the variables of interest are used to derive the parameters.

The populations we deal with in forest inventory can be described by a long list of attributes, some associated with the trees, others with the environment in which they grow or have grown. Information on several of these attributes may be available to the survey analyst at the time a survey is planned. Available information that is in some way associated with the attribute of interest for the survey can often be incorporated in the design for stratification or assignment of selection probabilities and in the estimators in the form of predictors. Common examples of auxiliary information in forest inventory include aerial photography, satellite imagery, and various thematic maps. As a rule, sampling designs/estimators that exploit auxiliary information optimally are more efficient than design/estimators that ignore this information. The most common sampling designs in the two groups are listed in Fig. 3.4.

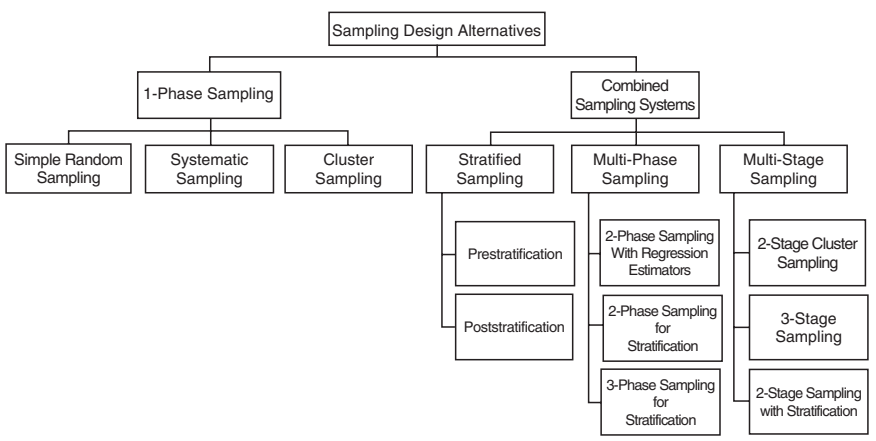


Fig. 3.4. Sampling design alternatives. The *boxes* on the *rightmost branch* list designs that incorporate auxiliary information in the design and/or the estimation phase. The *boxes* on the *leftmost branch* list designs that do not incorporate auxiliary information. (Courtesy of Pelz and Cunia 1985)

The following sections describe the major sampling designs currently used in forest inventories. The designs can apply to any population large or small, spatially contiguous, or spatially dispersed. Large forest enterprises may conduct several different inventories, each using a different design and estimators. Results from an inventory or several inventories may be used in postinventory analyses to provide estimators for specific subpopulations, updating, and forest modeling (van Deusen 1996; McRoberts et al. 2002; Tuominen et al. 2003). The utility of inventory results for these additional uses should be factored into the sampling strategy.

3.3.1 Simple Random Sampling

We begin the detailing of common inventory sampling designs with simple random sampling (SRS) not because it is particular widely used in its simplest form but because a presentation of SRS and its estimators will make it easier to comprehend and appreciate more complex designs and their estimators.

In finite-population SRS n units are selected at random from the N units comprising the entire population. The selections are done in such a way that all possible distinct samples of size n have the same selection probability. Since there are $\binom{N}{n} = \frac{N!}{n!(N-N)!}$ possible samples the selection probability of each (sample) becomes $1 \Big/ \binom{N}{n}$. The probability that a single unit/element is in the sample is n/N. The principle of equal sample inclusion probabilities extends naturally to infinite populations but we have, of course, no means of calculating these probabilities.

The simplest way of selecting units is to number all the elements of the population, to choose n numbers randomly, and to include the elements with the corresponding numbers in the sample. Here, however, it must be ensured that all the elements are listed – a circumstance that practically never occurs in forest inventories. In forest inventories using SRS, aerial photography, satellite imagery, or a map is needed to establish a frame from which the sample is to be taken. X/Y coordinates are randomly chosen and the survey is then conducted at the corresponding points. These coordinates may designate the centers of fixed-area plots, point samples, stem distance methods, or a fixed number of trees located nearest to the randomly selected coordinate whether it is sampled or not (Sect. 3.4).

There are two types of SRS: SRS with replacement and SRS without replacement. In SRS with replacement the same element may be drawn twice or more often and thus the elements are given the same selection probability at every draw, i.e., n/N for each draw in a finite population. In SRS without replacement a selected unit/element is removed from the sampling frame before the next

unit/element is selected. Thus, for a distinct element remaining in the sampling frame after completion of k draws the probability of selection at the $(k+1)$th draw is $(n–k)/N$ and so on for $k = 0, \ldots, n - 1$. Whenever a unit/element is selected more than once the sample will contain "copies" of the sample record associated with the unit. Copies of a sample record provide no new information about the population; hence, sampling with replacement is considered potentially wasteful and less efficient. The rationale for detailing with-replacement sampling is that some variance estimators can only be derived if we assume sampling with replacement (Brewer and Hanif 1983). These with-replacement estimators are then used as approximations to an estimator for sampling without replacement. When the sample sizes are small compared to the size of the population of interest the differences will often be trivial. In the following only SRS without replacement is considered.

3.3.1.1 Estimating the Population Mean

The population mean is given by

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Under SRS the sample mean $\hat{\overline{Y}}$ is an unbiased estimator of the population mean:

$$\hat{\overline{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

3.3.1.2 Sampling Error

In natural resource surveys the variable values associated with a sampling unit vary from unit to unit. The degree of variability depends on the variable and the population in question. Variability is thus an essential characteristic of survey sampling. The standard error and its square, the variance, are useful measures to quantify the variability or dispersion of values for individual population units about their mean. The variance of individual unit values of a variable, say y_i, is defined by

$$\mathrm{var}\left(y_i\right) = \frac{\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2}{N-1}.$$

The standard deviation of the population attribute is the square root of the variance:

$$SD\left(y_i\right) = \sqrt{var\left(y_i\right)}.$$

Under SRS, a sample-based design-unbiased estimator of the population variance is

$$\hat{\text{var}}\left(y_i\right) = \frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n-1}.$$

A sample-based estimator of the population standard deviation of y_i, $\hat{SD}\left(y_i\right)$, is obtained by taking the square root of $\hat{var}\left(y_i\right)$.

It is often convenient to remove the effect of the measurement scale from estimators of variability. Variances expressed in relative terms with respect to the mean of the variable to which they refer are scale-invariant. The coefficient of variation is a popular scale-invariant measure of variation:

$$\text{CV}\left(y_i\right) = \frac{SD\left(y_i\right)}{\overline{Y}}.$$

Sample-based estimators of the coefficients of variation are obtained by replacing the population quantities by their respective estimators.

The interunit variation means that the sample mean based on a sample of size n will also vary from one sample to the next if we repeat the sampling. Let us assume a population of size N, from which we take five samples each of size n. For each sample we calculate, say, the mean $\hat{\overline{Y}}_j$, $j=1, \ldots, 5$. Obviously, the five means will vary. Consequently any sample estimate of a parameter is subject to an error due to the randomness of the sample. This error is termed the standard error of sampling, or simply the standard error.

The larger the variability of the units, the larger is the standard error in separate estimates. Luckily, it is not necessary to take several samples from the same population in order to determine the standard error. We can make use of the *central limit theorem* which says that the mean of n randomly selected population values of a variable is asymptotically $(n, N \rightarrow \infty)$ normally distributed with a variance that is the variance of the random variable divided by n (Casella and Berger 2002). With SRS an estimator of the sampling variance of an estimate, say, $\hat{\overline{Y}}$ is

$$\hat{\text{var}}\left(\hat{\overline{Y}}\right) = \left(\frac{N-n}{N}\right)\frac{\hat{\text{var}}\left(y_i\right)}{n}.$$

An unbiased sample-based estimator is obtained by replacing the population variance var (y_i) by a sample-based estimate of this variance. The square root of the sampling variance is the standard error of sampling:

$$\widehat{\text{SE}}\left(\hat{\overline{Y}}\right) = \sqrt{\hat{\text{var}}\left(\hat{\overline{Y}}\right)}.$$

The quantity $(N–n)/n$ accounts for the changes in selection probability for sampling without replacement and is termed the *finite-population correction factor*. If the sampling fraction n/N is small, the finite-population correction factor will be close to 1. We get $\hat{\text{var}}\left(\hat{\overline{Y}}\right) \simeq \frac{1}{n}\,\hat{\text{var}}\left(y_i\right)$ when $n/N \simeq 1$. Omitting the

finite-population correction factor results in a slight overestimation of the true variance. For practical purposes, the finite-population correction factor needs not be considered if the sampling fraction is smaller than 5%. Note the previous formula extends naturally to population parameters other than the one chosen here.

3.3.1.3 Confidence Intervals for Sample Estimates

The concept of standard error is often not intuitively clear to many users of inventory data and they may find it difficult to assess the significance of a standard error and interpret it correctly. Estimates arising from a sample-based inventory ought to include a measure of uncertainty of the estimates. Confidence intervals for sample estimates provide an intuitive easily understood measure of the impact of a standard error.

A confidence interval for an estimate gives the range within which one can expect the true population parameter to be located. The bounds of the confidence interval are termed confidence limits. The interval should have the property that the probability of the true value being located within the confidence limits is known, say $1-\alpha$. The quantity $1-\alpha$ is called the confidence coefficient and the interval is called the $100(1-\alpha)\%$ standard interval. Typically the 95% confidence interval is presented ($\alpha=0.05$). The 95% confidence interval covers for 95 out of 100 replicate samples of size n the true value of the population. Conversely, there is a 5% chance that the true value is outside this interval. A specific sample-based estimate of the confidence interval either includes the true value or does not.

The distribution of sample estimates under repeat sampling is usually assumed to be normal (invoking *the central limit theorem*) with a mean equal to the estimate obtained and a variance equal to the estimated variance divided by the sample size. Under this assumption and continuing with the previous example with a population mean as the parameter of interest, the lower $\left(\hat{\bar{Y}}_{\mathrm{L}}\right)$ and upper $\left(\hat{\bar{Y}}_{\mathrm{U}}\right)$ limits of the $100(1-\alpha)\%$ confidence interval for a sample-based estimate are

$$\hat{\bar{Y}}_{\mathrm{L}} = \hat{\bar{Y}} - t_{n-1,1-\alpha/2} \times \widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right)$$

and

$$\hat{\bar{Y}}_{\mathrm{U}} = \hat{\bar{Y}} + t_{n-1,1-\alpha/2} \times \widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right).$$

$t_{n-1,q}$ is the qth quantile of Student's t distribution (Casella and Berger 2002). Values of t for $n=50$ and some common values of α are given in Table 3.2. For large sample sizes, say larger than 50, the quantiles of the t distribution are very close to the corresponding quantiles of a standard normal distribution z_q.

Table 3.2. Values of t for some common values of α (n=50)

	Confidence coefficient (%)				
	50	80	90	95	99
α	0.50	0.20	0.10	0.05	0.01
$t_{49,\ 1\ -\ \alpha/2}$	0.68	1.30	1.68	2.01	2.68
$t_{49,\ \alpha/2}$	–0.68	–1.30	–1.68	–2.01	–2.68

It is customary to use the standard normal quantile for large n. For large n and 95% confidence probability, t is approximately 2, and the confidence interval is called the 95% confidence interval. For a 68% confidence probability, t is approximately 1.

A word of caution is appropriate. Large sample sizes are usually needed to assure that a 100(1–α)% confidence interval has the desired properties. We say that the nominal coverage is asymptotically correct. Distributions of sample statistics obtained from small samples may be skewed and not well described be either Student's t distribution or the normal distribution. Confidence intervals obtained from such distributions may be liberal (they cover the true parameter below the nominal rate) or conservative (they cover the true parameters above the nominal rate). Various resampling schemes and improved approximations of the sampling distribution have been suggested to remedy problems of this nature (Davison and Hinkley 1988; Barndorff-Nielsen and Cox 1989; Shao 2003).

3.3.1.4 Estimating the Population Total

The population total, Y, is obtained by multiplying the population mean $\overline{Y}$ by the number of elements in the population:

$$Y = \sum_{i=1}^{N} y_i = N \times \overline{Y}.$$

An unbiased estimator of the population total $\hat{Y}$ is

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} y_i = N \times \overline{Y}.$$

As Y is N times the estimator $\bar{Y}$, an unbiased estimator of the variance of $\hat{Y}$, $\hat{\text{var}}\left(\hat{Y}\right)$, is

$$\hat{\text{var}}\left(\hat{Y}\right) = N^2 \times \hat{\text{var}}\left(\hat{\overline{Y}}\right).$$

The standard error of $\hat{Y}$ and the upper and lower confidence limits are

$$\widehat{\text{SE}}\left(\hat{Y}\right) = \sqrt{\hat{\text{var}}\left(\hat{Y}\right)} = N \times \sqrt{\hat{\text{var}}\left(\hat{\overline{Y}}\right)},$$

$$\hat{Y}_{\mathrm{L}} = \hat{Y} - t_{n-1,1-\alpha/2} \times \widehat{\mathrm{SE}}\left(\hat{Y}\right),$$

and

$$\hat{Y}_{\mathrm{U}} = \hat{Y} + t_{n-1,1-\alpha/2} \times \widehat{\mathrm{SE}}\left(\hat{Y}\right).$$

Note, in the above we assumed N is known without error. At times we do not know N but we may have an estimate of N, an estimate with a sampling error. To account for the added uncertainty, we must add an estimate of the error stemming from imperfect knowledge of N. The general technique for obtaining a variance estimator of an estimate that depends on several random variables is based on a Taylor series approximation (Kotz and Johnson 1988). The technique also goes under the name of the delta technique (Kendall et al. 1983). Specifically, let Y be a function f of a set of predictor variables, i.e., $Y = \mathrm{f}(X_1, X_2,...,X_v)$. A first-order Taylor series approximation of the variance of, say, a mean $\overline{Y}$ is

$$\mathrm{var}\left(\bar{Y}\right) \simeq \sum_{j=1}^{v} \left(\frac{df}{dX_j}\right)^2_{X_j = \overline{X}_j} \mathrm{var}\left(\overline{X}_j\right)$$

when the predictors $X_1, X_2,...,X_v$ are independent (no covariance). Approximations to variances of a sum, a ratio, or a proportion are found by straightforward extensions. At times the predictors $X_1, X_2,...,X_v$ will not be independent of each other. Under these circumstances the covariance of all possible pairs of predictors must be added to right-hand side of the previous equation.

3.3.1.5 Determining Sample Size

The sufficient SRS sample size is determined by the inherent (natural) variability of the attribute values of the population, the degree of precision required for the results, and the confidence coefficient we wish to apply to confidence intervals of sample estimates. In SRS the sample size needed to satisfy a desired precision (E) with a confidence coefficient of $100(1-\alpha)\%$ is calculated, for say a mean, according to

$$n = \frac{t^2_{n,1-\alpha/2} \times \mathrm{var}(y)}{E^2}.$$

In practice we would normally not know the variance of the variable of interest. It must be replaced by an estimate derived from historic information, related surveys, or simply from a qualified expert guess. Prudence dictates a conservative guess.

As the t value depends on the degrees of freedom, i.e., the sample size, calculations for small sample sizes must be done iteratively. During each iteration,

the n value determined in a previous iteration is used to determine the appropriate t value. Iterations are stopped when the upwardly rounded value of n no longer changes. Prodan (1965) suggested an alternative estimator of the variance derived from knowledge about the range of values y_i in the population (Beyer 1968). Provided that an estimate of the maximum and minimum values of y_i can be obtained, an approximate estimator of the variance is

$$var(Y) \simeq \left(\frac{\max(y_i) - \min(y_i)}{4} \right)^2 .$$

A usually very conservative "guess" is obtained by assuming that y_i is uniformly distributed between the maximum and the minimum. If this is indeed the case, the variance of y_i is $\left[\max(y_i) - \min(y_i)\right]^2 / 12$ (Snedecor and Cochran 1971).

3.3.1.6 Sampling for Proportions and Percentages

Some forest inventory results may be presented in terms of counts, proportions, or percentages. Examples would be tree count, proportion of burned forest, and percentage of teak volume in a forest. Counts, proportions, and percentages usually involve elements/units belonging to a defined class or exhibiting a given characteristic. Additional examples might include ownership types, proportion of trees with stem damage, or the percentage of the forest area that is difficult to access. When there are more than two mutually exclusive classes for an attribute/variable, the term "multinomial variable" is used. The results obtained on the basis of multinomial variables are presented as classifications on a nominal (e.g., tree species, soil type, access) or on an ordinal (e.g., stand layer, timber quality, burned status) scale. Multinomial variables are frequently analyzed and presented on the basis of a sequence of proportions or counts.

The following discussion is limited to the relatively simple case of SRS where each sampled unit exhibits a binary class value. There is extensive literature on the more complex analysis of proportions and percentages in other sampling designs (Kish 1965; Cochran 1977; Sukhatme et al. 1984; Agresti 1992; Lloyd 1999).

Binary variables assume one of two values, typically the value y_i=1 when the element/unit belongs to a given class and y_i=0 otherwise. The number of population elements in the class assigned a value of 1 is given by

$$Y = \sum_{i=1}^{N} y_i = N \times P,$$

where P is the proportion of elements/units in the population with a class value of 1. The proportion of elements that do not have the class attribute with the value of 1 is of course 1–P.

The population variance of y_i is

$$\operatorname{var}(y_i) = \frac{1}{N-1} \times \sum_{i=1}^{N} (y_i - P)^2 = \frac{N}{N-1} \times P \times (1-P).$$

Sample-based estimators of the population proportion and its variance are obtained from the previous equations after substituting n for N and adding carets to distinguish them from the true population values. When n is very large a good approximation to the standard error of $\hat{P}$ is given by

$$\widehat{\mathrm{SE}}(\hat{P}) \simeq \sqrt{\frac{\hat{P} \times (1-\hat{P})}{n-1}}.$$

The confidence interval for $\hat{P}$ can be calculated from

$$\hat{P} \pm \left(\left| t_{n-1,\alpha/2} \right| \times \sqrt{\frac{\hat{P}(1-\hat{P})}{n-1}} + \frac{1}{2n} \right),$$

where the last term is a correction for continuity, which is necessary as P is not a continuous variable. According to Cochran (1977) the omission of the continuity correction leads to confidence intervals which are too small (liberal). A detailed discussion of alternatives for the construction of confidence intervals for P is given by, for example, in Dees (1988) and Burk (1991). The standard normal approximation for binary confidence intervals is sufficient in situations, where $\hat{P}$ and n are not too small. Cochran (1977, p. 58) gives the smallest values for $n \times \hat{P}$ to which the normal approximation can still be applied. For example, for P=0.2 the smallest n is 200, while at P= 0.1 n should be larger than 600.

If more than one observation is made on the binary trait in every sampled unit the estimation of proportions has to be modified since the number of observations per sample unit can vary. Fixed-area plots are a typical example. The number of, say, trees per plot varies naturally between plots, The estimators of P and the variance of P are now given by

$$\hat{P} = \frac{\sum_{i}^{n} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i},$$

where subscript i refers to sample unit and subscript j to the population elements in the ith sample unit. There are m_i population elements in sample unit i of which a_i elements belong to the binary class given a value of 1. The corresponding estimator of variance becomes

$$\hat{\operatorname{var}}(\hat{P}) = \frac{1}{\overline{m}^2} \times \frac{\sum_i a_i^2 - 2\hat{P} \times \sum_i a_i m_i + \hat{P}^2 \sum_i m_i^2}{n-1},$$

where $\overline{m}$ is the mean number of population elements per sample unit.

Selection of an equal number of trees per plot as done in nearest-neighbor (NN) methods (Pielou 1970) can be treated as a special case where m_i is constant. In this case a_i is the area occupied by the selected trees (de Vries 1986).

The estimation of proportions when the attribute/parameter of interest has more than two classes can be treated as a special case of a binary class problem. In brief, one obtains simultaneously an estimate of the proportion for each class as if it was a binary class. Specifically, if the kth class out of K is considered, then elements/units in class k are given a value of 1 and all other elements/units a value of 0. The variance for each class is calculated as before for a binary class. In balanced samples the sum of the estimated proportions for the K classes will sum to 1, as they should. However, sample imbalances may cause a violation of the sum-to-1 constraint. Formulae for postestimation calibration of proportions to meet a sum-to-1 constraints and formulae for variance and covariance estimation for proportions are found in, for example, Wu (2003); Angers (1989); Agresti and Caffo (2000); Sison and Glaz (1995).

The estimation of confidence intervals for K classes has to take into account that estimates of precision are simultaneously given for K classes. Doing this requires us to distribute the global significance level α across the K interval estimates. Multinomial confidence intervals have been described by Gold (1963), Quesenberry and Hurst (1964), and Goodman (1964, 1965). The methods presented are based on the normal approximation (Fienberg and Holland 1973; Angers 1989; Sison and Glaz 1995) and they differ with respect to the approach to calculate simultaneous probability estimates. We limit ourselves to the presentation of the Bonferroni method (Miller 1981), which can be used for the adjustment of significance levels in multiple significance tests. The Bonferroni method distributes the global significance level equally across the individual classes; hence, the individual confidence coefficients become $1-\alpha/K$. For a simultaneous confidence coefficient of 0.95 ($\alpha=0.05$) the Bonferroni-adjusted significance level for each of four classes is $\alpha_4=0.05/4=0.0125$. The t value associated with this simultaneous Bonferroni-type confidence coefficient is obtained from tables or from one of the many statistical software programs available today.

Estimates of P are of course restricted to the interval from 0 to 1. At times an estimate of the lower bound of a confidence limit will be negative or the upper limit could exceed 1. A logistic transformation of P to $\log\left(\frac{P}{1-P}\right)$ and an estimation of the confidence limits on this transformed scale and a subsequent back transformation to the original scale resolves this type of problem since the back transform of a logistic variable is always between 0 and 1 (Lloyd 1999). For example, if we have an estimate of 0.1 for P and the standard error of this estimate is 0.0948 with nine degrees of freedom ($n=10$), the limits of our 95% standard interval would be -0.086 and 0.286. The logistic transform of 0.1 is 2.197 and the standard error on the logistic scale is found by application of the previously mentioned delta technique to be $1/\sqrt{n\times\hat{P}(1-\hat{P})}$ or 1.054. The standard interval on the logistic scale is therefore $(-4.263, -0.131)$. After a back transformation to the original scale the interval is (0.0139, 0.467).

3.3.1.7 Ratio Estimators

Ratio estimators are widely used in forest inventories. Any attribute related to an area such as, for example, the number of trees per hectare or the volume per hectare, is defined as a ratio of two different attributes. The population ratio R is obtained by dividing the population total of the attribute (total volume, total number of stems) in the numerator of the ratio by the population total of the attribute in the denominator of the ratio. A sample-based estimator of R is the ratio of the two sample estimates of population totals. This is a *ratio of means estimator*, which has a bias of the order of $1/n$. There is no unbiased sample-based estimator of R.

$$\hat{R}_{yx} = \frac{\sum y_i}{\sum x_i} = \frac{\hat{\bar{Y}}}{\hat{\bar{X}}},$$

where summation is over the sampled values. For sample sizes n over 30 the bias if often negligible, but skewed population distributions of Y and especially X can introduce a serious bias in a sample estimate (Hess and Bay 1997; Rao 1988). The variance of a ratio of means estimate is

$$\hat{\text{var}}\left(\hat{R}\right) = \frac{\hat{\text{var}}\left(y_i\right) + \hat{R} \times \hat{\text{var}}\left(x_i\right) - 2\hat{R} \times \hat{\text{cov}}\left(y_i, x_i\right)}{n \times \hat{\bar{X}}^2},$$

where $\text{cov}\left(y_i, x_i\right)$ is the covariance between the two attributes/variables y and x. The covariance is estimated as

$$\hat{\text{cov}}\left(y_i, x_i\right) = \frac{\sum y_i x_i - n^{-1} \sum y_i \sum x_i}{n-1},$$

where summation is over the sampled values.

It is often possible to estimate a plot-level ratios R_i for each plot in a sample. For example, the number of trees per hectare in each plot. However, the estimation of the population ratio should not be based on these individual ratios because the mean of these individual ratios as an estimate of R has more bias than the ratio of means even if n is large (Cochran 1977). The exception is when $y_i = R \times x_i$ for all elements in the population. In this case the mean of Y_i/X_i is obviously R everywhere. Furthermore, the per unit ratio is often unstable and exhibits a large variance and a very skewed sampling distribution.

3.3.1.8 Advantages and Disadvantages of SRS

Strict adherence to the principles of simple random selection guarantees unbiased and consistent estimates of population parameters and their standard errors. Yet there are often other sampling designs for which the expected

sampling error for a given sample size is lower than the sampling error expected under SRS. The ratio of expected sampling variance under a design, say $\mathfrak{I}$ to that of SRS is called the efficiency or design effect of $\mathfrak{I}$ (Särndal et al. 1992). Relative to a more efficient design a SRS requires a greater sample size for a given expected standard error. This usually also means that the cost for a forest inventory based on SRS would be higher than the costs incurred under a more efficient sampling design. Note, however, that the expected efficiency of a design relies on theoretical expectations. The survey planner has to obtain estimates of the expected sampling error under different competing designs and their cost implications before a rational choice can be made.

A SRS design often requires a surprisingly large investment in organization, checking, and location of the samples, investments that can be more time-consuming and therefore more expensive than for other, more efficient procedures. Also, through random selection an irregular spatial distribution of sample locations may result, so the population as a whole is not uniformly represented (Fig. 3.5). Although these outcomes are fully expected under the SRS design it is clearly unsatisfactory and perhaps even unacceptable to proceed with a sample that one suspects will yield estimates that are far from the true population values. For these reasons, the SRS design is commonly applied only to smaller homogeneous subpopulations as part of a more complex design. In general, a stratified sampling design with many homogeneous strata and just two samples per stratum offers the most efficient design (Royall 1998). Finally, a SRS design offers few opportunities for a postsampling correction/calibration to mitigate the negative impact of a "poor" sample.

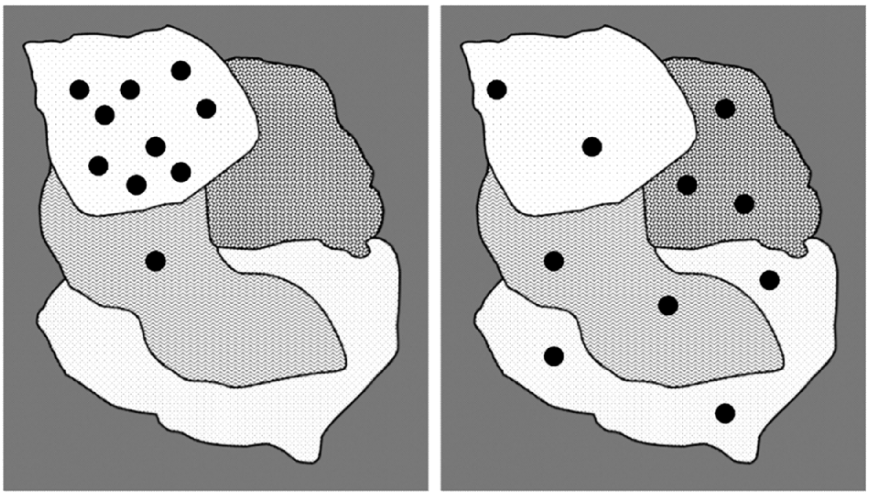


Fig. 3.5. Location of sample plots in two random samples

3.3.2 Systematic Sampling

As the term implies, the sample units are not randomly distributed across inventory area, but are drawn from the sample frame according to some systematic procedure. In systematic sampling the population is often subdivided into an exhaustive list of spatial units and n sample units are selected from this list by first choosing one unit at random, and then from this random starting position a systematic selection of the remaining n–1 units is made. Thus, the sample consists of one randomly selected unit. It is possible to derive unbiased estimators of totals and means from this sample but not of variance (Thompson 1992). The template for the spatial subdivision is often a regular grid of square cells or an equilateral triangular network. A major advantage of systematic sampling is that it is easy to locate the sample locations, the population is uniformly covered, and the efficiency is generally better than using SRS. As a rule, sample designs which are more "spatially balanced" will have a lower root-mean-square error when sampling from a population with patterned variation (Matérn 1980; Olsen et al. 1999; Stevens and Olsen 2004). As the joint selection probability of selecting two distinct units in the sample is either positive or zero, depending on the systematic sampling protocol, the selected elements are not independent of each other. This feature makes systematic sampling fundamentally different from SRS. Large-scale forest inventories, such as, for example, national inventories, often adopt a systematic design for the selection of sample location (Pelz and Cunia 1985; EC 1997)

In systematic sampling a sampling frame is constructed, which is a list of all sets of elements that are available for selection. When the basic sampling frame is in the form of a list (e.g., plant rows in a plantation) or consists of elements passing a certain point during a period of time (e.g., logs in a sawmill) the sample is generated by choosing elements from the frame that are separated by a constant interval L. If the population size N is a whole-number multiple of L then the sampling intensity n/N is equal to $1/L$ and all possible samples are size n. For example, the sampling frame for a plantation made up of 12 rows of plants could be the list [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12], where the number refers to the sequential position of a row of plants. For L=4 the sampling intensity is 1:4 and the sample will be composed of one of the following four possible subsets of rows [1, 5, 9], [2, 6, 10], [3, 7, 11], or [4, 8, 12]. In the sampling process one and only one of these four subsets would be selected as a sample.

When the basic sampling frame is in the form of a map, the universe is two-dimensional and we have several options for tesselating the population into a set of mutually exclusive and jointly exhaustive sets of units. Sometimes the tesselation is already provided to us in the form of suitable administrative units, like, for instance, counties or postal code districts. An example of a

simple geometric tesselation is in Fig. 3.6. The population displayed is completely contained inside a regular square grid of ten rows and ten columns. There are two spacing intervals to be determined for the sample selection, one for the rows, S_r, and one for the columns, S_c. The intervals chosen may or may not be equal. A random starting point for selecting the systematic sample is chosen by choosing at random one integer between 1 and S_r, say rS_r and one integer between 1 and S_c, say rS_c. Once found, the sample consists of the sample unit with coordinates $rS_r + i \times S_r$ and $rS_c + i \times S_c$ for $i = 0, \ldots, n-1$. For a square or rectangular sampling frame the sampling design is a 1 in $(S_r \times S_c)$ systematic sample. For $S_r = S_c = 5$ in a 10×10 grid (N=100), for example, this systematic sample design results in a sample size of 4. There are 25 possible samples of size 4. We only list five of them interleaving those samples that follow logically from a preceding sample. In the row–column notation given in Fig. 3.6 the possible samples of size 4 are [A1, A6, F1, F6], [A2, A7, F2, F7], . . ., [A5, A10, F5, F10], [B1, B6, G1, G6], . . ., and [E5, E10, K5, K10]. In this simple example all of the possible 25 subsets subset would have the same size n=4. An irregular spatial outline of the population or an irregular existing tesselation of the population may produce subsets of unequal size. When the sampling frame is not a square or a rectangle the discrepancy between the desired sampling intensity $1/(S_r \times S_c)$ and the actual sampling intensity could become large.

Systematic sampling can be implemented as a special form of either a simple cluster sampling or a two-stage cluster sampling. Implemented as a simple cluster sampling, each of the 25 possible subsets is considered as a single cluster and only one is selected (Fig. 3.7).

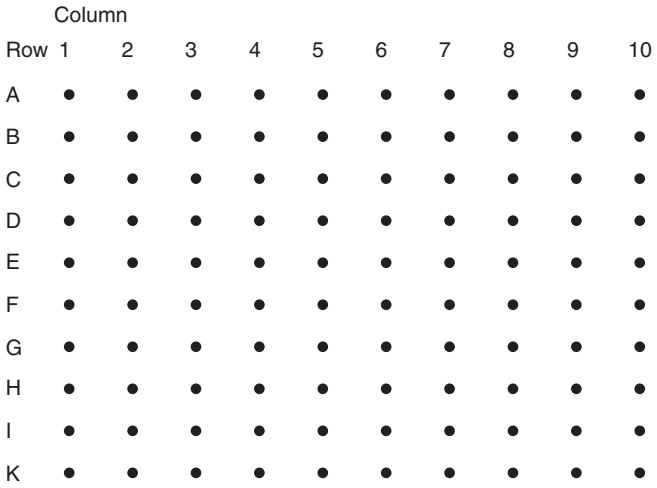


Fig. 3.6. A two-dimensional basic sampling frame

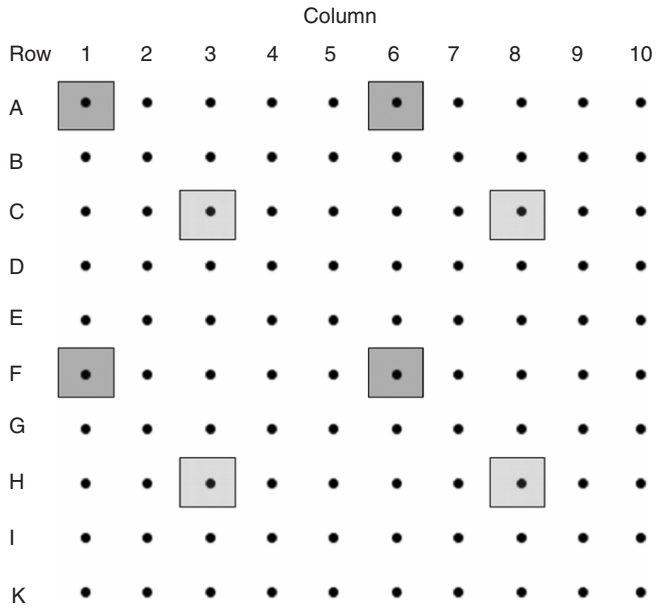


Fig. 3.7. Systematic sampling implemented as a cluster sampling. A cluster is composed of four shaded squares. The population contains 25 clusters, of which two are selected

When systematic sampling is implemented as a two-stage cluster sampling the sampling frame is divided into n compact clusters, each containing $S_r \times S_c$ elements. A subsample of size 1 is drawn from each of the n clusters. In the previous example n=4 compact clusters would be generated and one element selected from the 25 units inside each cluster (Fig. 3.8).

Estimators for the variance of the overall total and mean vary according to the way the systematic sampling is implemented (Cochran 1977). However, it is generally impossible to provide unbiased estimators of the variances when systematic sampling is used. Attempts have been undertaken to find estimators with little bias and low variance (Bellhouse 1985; Wolter 1985; Sherman 1996). The most commonly used approach is based on the assumption that a systematic sample is equivalent to a random sample; however, this assumption holds only when population attributes are randomly distributed over the population. With the assumption of SRS equivalency, means and totals are computed using the formulae applicable to SRS (Sect. 3.3.1). SRS estimators of standard errors applied to estimates from systematic sampling are usually conservative; they overestimate, on average, the actual error. An overestimation of about one third is not unusual, but more extreme results have been reported (Hartley 1966; Bellhouse 1988; Stehman 1992).

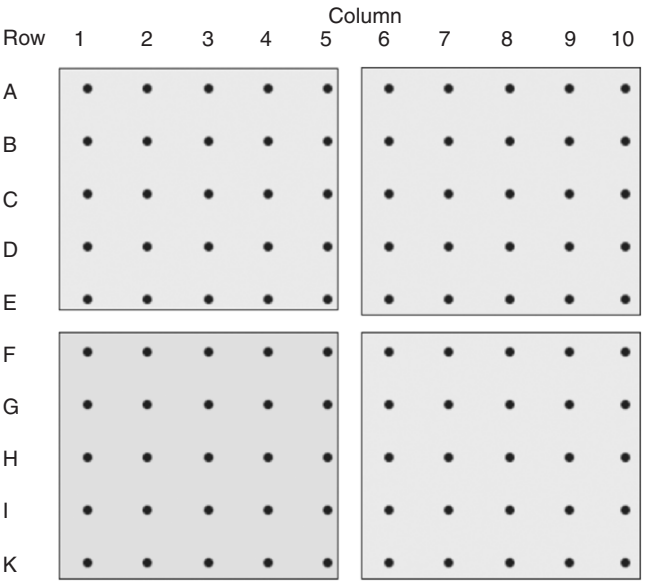


Fig. 3.8. Systematic sampling implemented as a two-stage cluster sampling

A series of variance estimators under systematic sampling have been proposed as more attractive than those flowing from the assumption of SRS. Jessen (1942), Yates (1949, 1981), and Cochran (1977) suggested a procedure that involves grouping of pairs of adjacent sample units (clusters). Each grouped pair of units is considered as a single stratum in a stratified sampling scheme, which allows the computation of a within-stratum variance, an estimate needed for the purpose of correcting the "inflated" SRS variance estimate. This procedure leads to corrected variance estimates that are biased, the bias being either positive or negative. Yates (1981) suggested for the situation of two-dimensional sampling to combine sample units into blocks of 2×2 units. Each block is then considered as a stratum in stratified random sampling. Instead of using squared residuals for computing a sampling variance, von Neumann et al. (1941) proposed using the sum of squared differences between adjacent sample units. This method is known as the method of squared differences (Ekström and Sjöstedt de Luna 2004; Stevens and Olsen 2004). The ideas of Neumann and Yates have been developed further in the form of NN estimators of local expectations and estimating the variance from the squared differences between the sample values and NN "predictions" (Ekström and Sjöstedt de Luna 2004; Stevens and Olsen 2004).

In systematic sampling the sample size is usually determined on the basis of optimizing a SRS design, which, for the reasons outlined earlier regarding the conservative nature of SRS estimators, eventually leads to somewhat more accurate inventory results than originally anticipated. In a systematic sampling from a square grid the sample size determines the scale of the grid to be used for implementing the systematic selection process. For a population occupying a square with an area F and a desired sample size of n, the appropriate grid spacing between sample locations should be $\sqrt{F/n}$. It gets a bit more complicated for population with an irregular outline, but in principle one finds the smallest rectangle that completely contains the population and then finds the grid spacing as before. Since some sample units may fall completely outside the population, the sample size is increased by trial and error until a satisfactory solution is obtained. For those who decide to use a triangular grid as the frame for a systematic sampling process, the distance between points becomes $\sqrt{2/3 \times F/n} \simeq 1.075 \times \sqrt{F/n}$ when the population occupies a square. In a triangular grid the distance between rows and columns is $0.557 \times \sqrt{F/n}$. To facilitate location of the sample units in the field, it is common practice to round off distances to the nearest 0.1 m. In establishing the grid, however, it should be borne in mind that some points may fall outside the target population. The density of the grid should be increased according to the proportions of sample units expected to fall outside the population of interest.

Semisystematic sampling designs have been suggested as a compromise between a systematic sampling design and SRS. In a semisystematic design the selection of spatially close sampling units is made less likely than by SRS or the number of selections from a single row, or column, is constrained to help achieve a more representative sample. Cox et al. (1997), Stein and Ettema (2003), and Stevens and Olsen (2004) provide examples of these semisystematic designs.

3.3.3 Cluster Sampling

Every sampling design is based on the division of the population into clearly defined units. The smallest units into which a population can be divided and which can be used for sample selection are the elements. Forest stands, lakes, road segments, and geometrically defined spatial polygons are but a few examples of units. Trees, snags, orchids, and deer are examples of elements. In cluster sampling two or more elements or two or more units are included in the sample at each sample location. The inclusion of two or more units/elements at each sample location intensifies the sampling effort at each sample location. The cost of including more than one unit/element at each single sample location is often

modest compared with the cost associated with travel and measurement of one unit/element per sample location. One example for such clustering is the surveying of trees (elements) on sample plots (cluster/unit). Another example is the establishment of three or four sample plots in a fixed geometric configuration at each sample location instead of just a single plot. The grouping of elements into clusters lent the procedure its name.

From the viewpoint of expense, such grouping of sampled elements/units is justifiable. However, adding two or more "extra" units/elements at each sample location does not necessarily mean that the sample size can be reduced by a factor equal to the number of added elements/units at each sample location. The trade-off between the number of elements/units sampled per sample location and the number of sample locations to visit in order to achieve a target precision on estimates of population parameters depends on the distribution of the variance of attribute values across spatial and temporal scales. In general, the less the variance of the attribute value is within a cluster, the less is the efficiency of clustering sample observations relative to a SRS of individual units/elements. This is intuitively clear. If units/elements in a cluster are more alike than units/elements selected at random then we do not learn as much about the population from one cluster with m units/elements as we would learn from m independent units/elements. Thus, for a cluster sampling approach to be attractive in terms of precision for a given overall number of sampled units/elements, the variation within a cluster must be large relative to the among-cluster variance.

In forests the variation in tree attribute within sample plots of, say, 100 m^2 is often as large as the variance between such plots (Correll and Cellier 1987; Saborowski and Smelko 1998; Barnett and Stohlgren 2003; Gray 2003). We can exploit this large small-scale variation by adopting a cluster sampling design for our forest inventories. In most forest surveys, the use of clusters with several elements (more then 10) and three to four plots is often fully justified in terms of both cost and overall precision. Conversely, in homogenous forest areas a cluster composed of more than a few elements per sample location would be inefficient.

An efficient cluster sampling design offers an attractive balance between the cluster size (m) and the number of sample locations to visit (n). The balance is a function of the spatial distribution of attribute values. The survey designer must have, at least, a working knowledge of how this distribution will affect the efficiency of cluster sampling with different cluster sizes and different spatial configurations of elements/units in a cluster (Smith 1938; Kleinn 1996; Magnussen 2001).

Cluster sampling is common in forest inventory. Examples are the national forest inventories of Sweden, Finland, Austria, France, the USA, and Germany (Köhl 1990). Clusters are square, rectangular, or have more complex shapes. Sample plots are placed in a fixed geometric configuration within each cluster.

The term cluster is rarely used explicitly for the spatially grouped plots. Instead the word tract has become widely accepted as a quasi-synonym for a cluster (of sample plots). A classic example of cluster sampling is provided by the Camp Unit System(Fig. 3.9), introduced in Thailand for inventorying teak stands (Loetsch 1957). A camp located in the center of a cluster is surrounded by what is termed satellites, each satellite comprising several sample plots that can be surveyed by a field team within a single day.

The simplest form of cluster sampling is the survey with clusters of constant size. To facilitate the understanding of cluster sampling this version of cluster sampling is detailed. Note, however, that cluster size in forest surveys is rarely constant. Fixed-area sample plots are in effect clusters of trees. It is obvious that the cluster size, i.e., the number of trees per plot, changes from cluster to cluster. Even clusters (tracts) designed with a fixed number of plots in a fixed geometrical configuration usually have no sample data for units/elements that are outside or straddle the population boundary.

The sample-based estimator of the population mean for one-stage cluster sampling with n clusters of equal size (m) or equal numbers of sample units per cluster selected by SRS is

$$\hat{\bar{Y}}_{\text{clust}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\hat{\bar{Y}}_i,$$

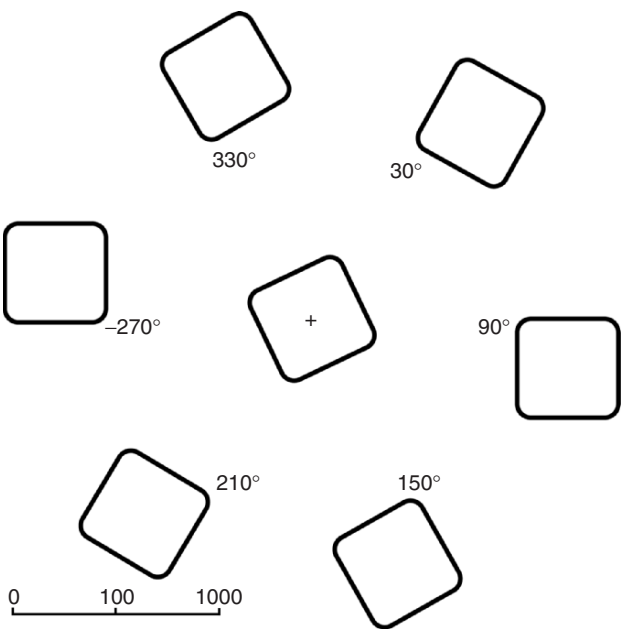


Fig.3.9. Camp Unit System

which is simply the average of cluster averages. The corresponding estimator, for sampling with clusters of unequal size, is a weighted average of cluster means with weights proportional to cluster size (Cochran 1977). The estimator of the sampling variance of the estimated population mean in a finite population composed of N clusters is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{\text{clust}}\right) = \frac{\left(1 - \frac{n}{N}\right)}{n-1} \sum_{i=1}^{n} \left(\hat{\bar{Y}}_i - \hat{\bar{Y}}_{\text{clust}}\right)^2$$

or simply the among-cluster variance of cluster means. Estimators for population totals are obtained by dropping the averaging of within-cluster observations.

When sample clusters are located at random locations or when the population is not tessellated uniquely by the clusters, the finite-population correction factor is dropped from the variance estimator. Many cluster shapes (tracts) used in forest inventory do not tessellate the population completely. The tessellation would produce overlaps of clusters or leave gaps between clusters. The selection probability of each cluster would no longer be equal and joint selection probabilities would have to be calculated for every pair of possible clusters. Unequal cluster sizes would change the variance estimator to account for unequal weighting of the squared deviations from each cluster. Again, the weight would be proportional to cluster size.

Estimators for clusters of unequal size are described in Cochran (1977) and Sukhatme et al. (1984). The optimization of forest inventory cluster sampling designs with clustering of fixed-area and variable radius sample plots is discussed by Scott (1981) and Köhl and Scott (Scott 1994).

3.3.3.1 Two-Stage Cluster Sampling

In two-stage cluster sampling, or simply two-stage sampling, the entire population is divided into N clusters. A sample of n clusters is selected. The ith cluster is assumed to be subdivided into M_i equally large smaller units called elements or simply second-stage units. A sample of m_i second-stage units is taken from the ith cluster. Thus, the sample is taken in two steps: first n clusters are selected, then a sample of secondary units is taken from each cluster. In a forest sampling context the clusters could be, for example, 1-km^2 Advanced Very High Resolution Radiometer (AVHRR) pixels and the secondary unit could be a 25×25 m^2 Landsat Enhanced Thematic Mapper Plus (ETM+) pixel, and if selected the attributes of trees in this pixel would be measured by some field procedure. A forest stand can also be viewed as a cluster and the inventory plots placed within this stand acting as secondary units. Two-stage cluster sampling is a very flexible design and applies well to a variety of applications. At

each stage a different selection scheme can be applied; this includes stratified selection of clusters and selection of secondary units chosen with equal probability, and selection of the secondary units with probability proportional to size (PPS)/prediction.

Two-stage procedures are generally preferable where localized systematic trends are expected. Systematic trends may occur in mountainous regions, along rivers and water bodies, and where natural history or anthropogenic effects have shaped vegetation and land use into distinct mosaics. The first-stage sampling then primarily serves to isolate the systematic variation to the first-stage units.

Two-stage sampling designs are frequently employed in forest inventory (Fig. 3.10). In its simplest incarnation every cluster contains the same number of secondary units, and both clusters and secondary units within clusters are randomly selected at each stage. Two-stage sampling is particularly attractive

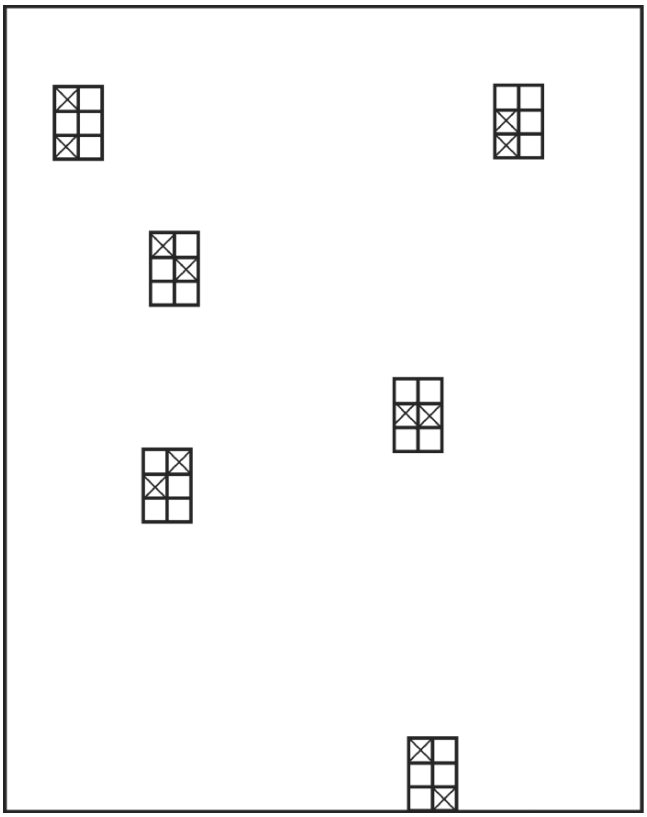


Fig.3.10. Two-stage cluster sampling

when access to individual sample locations is time-consuming and costly; a topical situation in many tropical forest regions.

Two-stage sampling is distinctly different from two-phase sampling. In the latter, an auxiliary variable is sampled and measured in the first phase and a target attribute/variable is measured in the second phase on a small subset of the first-phase sample units. The association between the auxiliary and target attribute is exploited, usually via linear regression, to allow a prediction of the target attribute for all those units/elements sampled in the first phase for which only the auxiliary attribute/variable was recorded. In two-stage sampling the stratification of the population into clusters and second-stage units serves only as a conduit for determining selection probabilities.

Aerial photography, for example, can be used in both design types. In a two-phase sampling design the photographic images would be used to estimate, by some form of interpretation, the value of an auxiliary attribute for a series of sample units (stands, plots). The attribute of interest would be measured by standard field procedures on a small subset of these units. In a two-stage design, however, the photographic images would only serve to stratify the population into clusters (stands, tracts) and then a number of clusters would be selected at random, after which a number of secondary units would be selected from each cluster for recording of the attribute of interest. If a full coverage of the inventory area by aerial photographs is not possible, then a modification of the two-stage estimators given next is required (Saborowski 1990).

With SRS sampling of n clusters out of N population clusters and sampling of m_i secondary units within the ith cluster composed of M_i secondary units the unbiased two-stage estimator of the population total is

$$\hat{Y}_{\text{clust2}} = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}$$

and the estimator for the population average per first-stage unit is obtained by division by N. We see that each second-stage sample is scaled to an unbiased estimate of the total in the sampled first-stage unit. For clusters of equal size and equal second-stage sample sizes the estimator of the total is greatly simplified. We leave the simplification as an exercise for our readers. The sample-based estimator of variance of the estimated total is

$$\hat{\text{v}}\text{ar}\left(\hat{Y}_{\text{clust2}}\right) = \frac{N^2\left(1 - \frac{n}{N}\right)}{n(n-1)} \sum_{i=1}^{n} \left(\hat{Y}_i - \hat{Y}_{\text{clust2}}\right)^2 + \frac{N}{n} \sum_{i=1}^{n} \frac{M_i^2}{m_i}\left(1 - \frac{m_i}{M_i}\right) \hat{v}ar\left(y_{ij}\right).$$

The two-stage estimator of variance is simply the variance of the first-stage totals plus the average of the second-stage variances scaled to the first-stage expectations.

Cochran (1977) provides a mnemonic way to construct variance estimators for multi-stage sampling. Simply put, the expected value of a population parameter θ in a *k*-stage sampling is $E_1\left(E_2\left\{\left[\ldots E_k(\theta)\right]\ldots\right\}\right), where E_l,\ l = 1,\ldots,k$ is the expectation of θ over all possible samples at the *l*th stage of sampling with all units at higher levels fixed. The expected variance of this expectation is $\mathrm{var}_1\left(E_2\left\{\ldots\left[E_k(\theta)\right]\ldots\right\}\right) + E_1\left\{\mathrm{var}_2\left[E_3(\theta)\right]\right\} + E_1\left(\left\{\cdots E_{k-1}\left[var_k(\theta)\right]\ldots\right\}\right)$, where $\mathrm{var}_l(\theta)$ is the *l*th-stage variance of θ with all higher stages fixed.

When all secondary units in all clusters are sampled (m_i=M_i for all *i*) we revert to estimators appropriate for single-stage cluster sampling. For *n*=*N*, that is all clusters are sampled, but m_i<M_i for at least some *i*, the two-stage estimator is identical to the estimator for stratified random sampling with clusters acting as strata.

At times we may have interest in estimating the attribute mean per first-stage cluster (unit). We can get the estimator for this average by dividing the estimator of the total by *N*. Similarly, the estimator for the variance of this mean is the previous variance estimator of the total divided by N^2. For large *N* the first-stage sample fraction *n/N* is negligible and can be set to zero without incurring more than a trivial bias in the resulting variance estimator. With small first-stage sample fractions, and equal secondary sample sizes in each cluster, the two-stage variance estimator for the mean simplifies to the variance of the first-stage (cluster) mean values divided by *n*. In this specific situation, the variance can be estimated from knowledge of first-stage cluster means only, a useful result for two-stage designs with systematic subsampling of second-stage units since we would not have an unbiased estimator of the second-stage variance.

Two-stage sample estimators of totals (mean) and variance for designs with unequal selection probabilities of first- and second-stage units have been developed (Mahalanobis 1946; Bowden 1979; Cochran 1977; Nusser et al. 1998).

The previous two-stage estimators assumed that the population was divided into a unique set of *N* clusters that, in turn, were subdivided into a fixed number of secondary units. *N* would be known in this situation. When first-stage clusters are merely a fixed-area-sampling device located at random in the population *N* is unknown and must be estimated by dividing the area of the population by the area of a first-stage unit. Also, when and a small number of inventory plots are placed at random or in some geometric configuration inside a first-stage unit we do not a priori know *M* but we can estimate *M* by dividing the area of the first-stage unit (cluster) by the area of an inventory plot. Possible errors in estimates of *N* and *M* must accounted for by extending the previous variance estimator to include these potential sources of variation.

3.3.3.2 Two-Stage Cluster Sampling for the Estimation of Proportions

When the sampling objective is to estimate the proportion P of secondary units with a specific attribute class value we are dealing with sampling for the estimation of a proportion. To facilitate the derivation of estimators of means and variance we define the attribute value (y_{ij}) of the jth secondary unit in the ith primary unit as 1 when the unit has the class value of interest, and 0 otherwise. Under SRS of n primary units and m_i secondary units out of a total of M_i units in the ith primary unit, the estimators of P and the sampling variance are

$$\hat{P}_{2\text{st}|\text{SRS}} = \frac{1}{n} \times \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{y_{ij}}{m_i}$$

and

$$\hat{\text{var}}\left(\hat{P}_{2\text{st}|\text{SRS}}\right) = \frac{1 - \frac{n}{N}}{n(n-1)} \text{var}\left(\hat{P}_i\right) + \frac{1}{n \times \sum_{i=1}^{n} m_i^2} \sum_{i=1}^{n} m_i^2 \hat{P}_i(1 - \hat{P}_i)\left(1 - \frac{m_i}{M_i}\right),$$

respectively, where $\hat{P}_i$ is the estimate of the proportion P in the ith primary unit. These formulae extend to cases with more than two classes by simply replacing an estimate of P by an estimate of the vector of class-specific population proportions.

3.3.3.3 Two-Stage Cluster Sampling with Stratification of the Primary Units

When the first stage is a sample of units in an aerial photograph or in a remotely sensed image we can often associate each first-stage unit with a specific land use or vegetation cover-type class. Within each first-stage unit we may sample one or more second-stage units. Second-stage units could be, for example, conventional forest inventory ground plots. Two-stage designs of this type are important for forest resource inventory. The first-stage units of a given stratum, say, h, are usually assumed to have a constant size and shape, but size and shape may vary from stratum to stratum. We shall consider a population with $N = \sum_h N_h$ first-stage units and M_h second-stage units in each first-stage unit of stratum h.

The estimator of the population mean per second-stage unit under SRS of n_h first-stage units and m_h second-stage units within each first-stage unit in stratum h is

$$\hat{\bar{Y}}_{2\text{st}|\text{STR}} = \frac{\sum_h \frac{N_h \times M_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{m_h} \sum_{j=1}^{m_h} y_{hij}}{\sum_h N_h \times M_h} = \sum_h W_h \times \hat{\bar{Y}}_h,$$

where w_h is the relative size of stratum h in terms of second-stage units. For unknown N_h and M_h, w_h is replaced by n_h/n (Cochran 1977, p. 328). With SRS in both stages an unbiased estimator of the variance of $\hat{\bar{Y}}_{2\text{st}|\text{STR}}$ is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{st}|\text{STR}}\right) = \sum\nolimits_h W_h^2 \times \left[\text{var}\left(\hat{\bar{Y}}_h\right)\left(1 - \frac{n_h}{N_h}\right) + \frac{\left(1 - \frac{n_h}{N_h}\right)\left(1 - \frac{m_h}{M_h}\right)}{n_h}\hat{\bar{\text{v}}}\text{ar}\left(\hat{\bar{Y}}_{hi}\right)\right],$$

where the term $\hat{\text{v}}\text{ar}\left(\hat{\bar{Y}}_{hi}\right)$ is the average weighted within first stage unit (cluster) variance with weights proportional to the number of second-stage units selected within a first-stage unit. Estimators for the population total and its sampling variance are obtained from the previous estimators by scaling to the number of second-stage units in the population.

Note, if the secondary units are drawn systematically from within the primary units, the design is not a true two-stage cluster sampling. In effect, the appropriate estimators to use in this case would be those given for single-stage cluster sampling.

3.3.4 Stratified Sampling

In stratified sampling we use auxiliary information to stratify the entire population into, say, H strata. Stratification aims at forming groups of elements (units) with more or less similar attribute values. The ideal stratification eliminates the within-stratum variation, hardly possible in practice. In the ideal case a single sample from each stratum would suffice to gain complete knowledge about the population parameter of interest since there is no within-stratum variance, all elements (units) would have the same attribute value. As we move away from the ideal case more samples are needed to precisely estimate the mean (total) of a stratum. In other words, stratification aims at dividing a population into a number of parts which are as homogeneous as possible.

Besides improving the variance efficiency of estimators other reasons to choose a stratified sampling design are (1) estimates for homogeneous subpopulations (strata) may be required, (2) the desired precision is not the same for all subpopulations, (3) assessment cost and/or attributes of interest are not the same for all subpopulations, and (4) different sampling protocols apply to different subpopulations.

A diverse spectrum of criteria can be used to stratify a population. Some examples are major timber type, vegetation type, stand structure, species mixtures, site quality, protective status, habitat, ecological sensitivity, wetland status, recreational use, nontimber resource values, and political and administrative

units. Where satellite imagery provides the auxiliary information the stratification is often done on the basis of the value of various indicators of vegetation types, such as, for example, the normalized-difference vegetation index (Sims and Gamon 2002; Wulder et al. 1996; Carlson and Ripley 1997; Gholz et al. 1997; Ricotta et al. 1999) or tessellated cap indices (Gustavson and Parker 1992; Bettinger et al. 1996).

Cochran (1977) and Dalenius and Gurney (1951) give generally valid rules for an optimum stratification of a single population parameter of interest when the distribution of an attribute value (y) or some auxiliary variable (x) related to y via a linear model is used directly for stratification. Given assumptions about the distribution of y and the relationship between x and y they develop an optimal solution for the number of strata and the selection of strata boundaries. The solution is optimal in terms of a minimum variance of the estimator of a population total (mean). For many typical distribution models for y, Gaussian, rectangular, triangular, and exponential, some five to ten strata appear to give substantial reductions in variance.

Once the criterion upon which to base the stratification has been decided, the inventory designer needs to consider the allocation of sample sizes to the strata. Again, both Dalenius and Cochran provide solutions for fixed total sample sizes or conversely for a fixed total cost of the inventory. When both stratification and sample size allocation are considered simultaneously the optimum design is often to maximize the number of strata and then take two samples per stratum. When the attribute values within a population are clustered, spatially or temporally, it is often good practice to define strata for each cluster.

Elements within a stratum are selected independently from the selections of elements in other strata. This accommodates stratum-specific sample sizes, selection criteria, and survey methods. When SRS is applied in all strata, the procedure is termed stratified random sampling.

In assessment, the strata are first evaluated separately and the results are then compiled to give overall estimates. The fact that stratified sampling renders it possible to compute estimations for subpopulations together with their precision is a distinct advantage.

In most practical situations auxiliary information suitable for a stratification of the population is readily available when planning for an inventory begins. By using this auxiliary information the efficiency of the inventory estimates of population parameters can be greatly improved. On the basis of the auxiliary information the population is divided into H distinct nonoverlapping strata or conversely the population is divided into N distinct nonoverlapping units and each unit is assigned to one and one only stratum. The strata cover the whole population without overlap, i.e., $N = \sum_h N_h$.

3.3.4.1 Sample Allocation

Deciding the number of samples to take from each stratum is perhaps the most important decision the inventory designer has to make once the decision to adopt a stratified sampling design has been made. Allocation of samples to strata may be done in various ways. The decision is often one of allocating a fixed total sample of size n to individual strata. The expected precision and cost of the resulting design can then be approximated from subject knowledge, experience, or qualified guesses. Design alternatives for different n can then be compared and the one judged most attractive against a set of global objectives is then favored.

One simple solution to the allocation problem prescribes an equal number of samples to be taken in each stratum. The sample size in stratum h is thus $n_h = \lceil n/H \rceil$, where H is the number of strata and $\lceil n/H \rceil$ denotes the smallest integer larger than n/H. Equal strata sample sizes, however, are seldom effective, as small strata are sampled with an disproportionately higher intensity than a large stratum.

A popular allocation scheme is the allocation of samples in proportion to the size of the strata. The size of stratum h (N_h) is measured in the number of elements (units). With this approach the sample size in stratum h becomes

$$n_h = N_h \times n \times N^{-1} \text{ with } N = \sum_h N_h .$$

At times the inventory designer will have some ideas or estimates of the expected stratum-specific variance of the attribute of interest. When both the within-stratum variance and the stratum size are considered together in the allocation problem and the objective is to minimize the expected variance of an estimate of a population total (mean), the solution is termed optimal allocation. Cost constraints may of course necessitate a shift away from this optimum towards an affordable design. The optimal allocation or Neymann allocation (Cochran 1977, p. 99) becomes

$$n_h = \left\lceil \frac{N_h \times \ddot{\sigma}_h}{\sum_h N_h \times \ddot{\sigma}_h} \right\rceil ,$$

where $\ddot{\sigma}_h$ denotes an a priori estimate of the standard deviation of the attribute of interest in stratum h.

Alternatives to the criteria of these allocation schemes include consideration of differences in costs for different strata and various survey methods. The overall importance of a stratum may further modify the allocation. Imposing limits on the minimum and maximum sample size in each stratum is also a popular scheme akin to a "minimax" strategy (minimize the risk of an extreme low precision; Amrhein 1995).

3.3.4.2
Estimation of Population Means and Totals Under Stratified Sampling

The estimators of the mean and the variance for stratum, say h, are as follows:

$$\hat{\bar{Y}}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$$

and

$$\hat{\text{var}}\left(y_{hi}\right) = \frac{1}{n_n - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \hat{\bar{Y}}_h\right)^2.$$

The estimator for the population means under stratified random sampling is

$$\hat{\bar{Y}}_{\text{STR}} = \sum_h \frac{N_h}{N} \times \hat{\bar{Y}}_h = \sum_h W_h \times \hat{\bar{Y}}_h.$$

Often we know neither the total population size N nor the size of individual strata (N_h, h=1, ..., H). If we then replace the stratum weights W_h by the sample-based weights $W_h = n_h/n$ we obtain a biased estimate. The bias remains constant as the sample size increases. When the attribute of interest is expressed in units per unit area, the area of the population (A) and the area of individual strata (A_h) is used instead of N, or N_h. Area-based strata weights then replace the weights based on size in units, or elements. For a stratum area A_h and total area $A = \sum_h A_h$ the area weight for stratum h becomes $wa_h = A_h \times A^{-1}$. When strata areas are known to within a negligible error, a situation that is common when the strata information comes from a classified remotely sensed image, the bias arising from using estimated weights $\widehat{wa_h}$ in place of the true area weights wa_h can safely be ignored. If proportional allocation is used and W_h is replaced by the area proportion of stratum h most of the potential gain of stratification compared with SRS is nevertheless retained.

The estimator for the variance of $\hat{\bar{Y}}_{\text{STR}}$ is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{\text{STR}}\right) = \sum_h \frac{W_h^2 \times \hat{var}\left(y_h\right)}{n_h}\left(1 - \frac{n_h}{N_h}\right).$$

The variance estimator is simplified if the sample fraction in each stratum is negligible and if the sample allocation is proportional to stratum size (area). We leave the simplifications as an exercise for the interested reader.

Estimates of population totals and variance of population totals are obtained from the previous estimators of the population mean and variance by multiplying the former by N and the latter by N^2.

3.3.4.3
Estimation of Proportions Under Stratified Random Sampling

For attributes on a nominal or an ordinal scale it is often desired to give their proportion within a stratum. Let $P_{h.c}$ be the proportion of sample units in stratum

h with attribute class value c in the sample from the hth stratum. An estimator of $P_{h.c}$ is

$$P_{h.c} = \frac{1}{n_h} \sum_{i=1}^{n_h} \delta_{hi.c},$$

where $\delta_{hi.c}$ is an indicator variable taking the value 1 if the ith sample in the hth stratum has attribute class value c and zero otherwise. The estimate of the proportion in the population is

$$\hat{\bar{P}}_{\text{STR.c}} = \sum_h \frac{n_h}{n} \times \hat{\bar{P}}_{h.c}.$$

With proportional allocation, and assuming that the stratum-specific final population correction factors can be ignored, the variance estimator for the estimated population proportion is

$$\hat{\text{v}}\text{ar}\left(\hat{P}_{\text{STR.c}}\right) = \sum_h w_h^2 \frac{\hat{P}_{\text{STR.c}}\left(1 - \hat{P}_{\text{STR.c}}\right)}{n_h - 1}.$$

Again, area weights (known or estimated) could also be used in place of sample size weights.

3.3.4.4 Design Effect

In many situations the surveyor will have a choice of sample design. We have already mentioned how variance efficiency, costs, and other practical considerations play a role in the final choice. A survey planner will often lack one or more critical pieces of information needed to make a truly optimal choice. Instead of optimizing a design it may be informative to know what variance to expect under a given design and how this variance compares with the variance of a "benchmark" design. The benchmark design is commonly the SRS design. The ratio of expected variance under the candidate design and the expected variance under the benchmark design is called the design effect, or DEFF for short (Kish 1965; Cochran 1977; Särndal et al. 1992). The candidate design is favored when the design effect is less than 1.

The design effect of stratified sampling with proportional allocation and SRS as the benchmark is

$$\widehat{\text{DEFF}}_{\text{STR: SRS}} = 1 - \frac{(n-1)\sum_h W_h\left(\hat{\bar{Y}} - \hat{\bar{Y}}_{\text{STR}}\right)^2}{\sum_h \sum_{i=1}^{n_h}\left(y_{hi} - \bar{Y}_{\text{STR}}\right)^2}.$$

We gather from this expression that the design effect is below 1 when the among-strata variance is made relatively large compared with the total variance. As the second term in the equation for the design effect is positive and less than 1, the design effect for a proportionate stratified design will always be less than 1, i.e., stratified random sampling with proportional allocation of

samples to strata is always more efficient than SRS. To the extent that the stratum means differ from each other, the second term will increase with a corresponding decrease in the design effect. Conversely as the among-strata variance increases the within-stratum homogeneity is on the increase. In conclusion then, stratified random sampling with proportional allocation of samples to strata may produce a significant decrease in the sampling variance relative to the variance expected for a SRS design with the same total sample size.

3.3.4.5 Poststratification

The term poststratification applies to a procedure for which SRS samples are stratified to a set of known strata after completion of the sampling. In other words, the auxiliary strata information was not used during the sampling process. Poststratification may apply to a field survey completed before a remote-sensing-based stratification becomes available. Poststratification facilitates forest surveys, as field sampling and analysis or interpretation of remote-sensing data can be done independently.

In its simplest form poststratification applies to data from a SRS. Using the previous notation for stratified random sampling but with the addition of ".ps" to distinguish a poststratum from an a priori stratum. The poststratification estimator of the population mean is

$$\hat{\bar{Y}}_{\text{STR.ps}} = \sum_{h.\text{ps}} W_{h.\text{ps}} \times \hat{\bar{Y}}_{h.\text{ps}}$$

and, assuming we can ignore the finite-population correction factor, the estimator of the sampling variance of the poststratified estimate of the population mean is

$$\hat{\text{v}}\text{ar}\left(\hat{\bar{Y}}_{\text{STR.ps}}\right) =$$

$$\frac{1}{n}\left[\sum_{h.\text{ps}} W_{h.\text{ps}} \times \hat{\text{v}}\text{ar}\left(y_{h.\text{ps}}\right) + \sum_{h.\text{ps}} W_{h.\text{ps}}\left(1 - W_{h.\text{ps}}\right) \times \frac{1}{n_{h.\text{ps}}} \hat{\text{v}}\text{ar}\left(y_{h.\text{ps}}\right)\right].$$

The first term in the variance estimator is identical to the variance under a stratified sample with proportional allocation of sample sizes and within-stratum SRS. The second term reflects an increase in the variance due to the random nature of the strata weights. The strata weights in poststratification are the expected value of a random binary variable taking the value of 1 if a sample is in stratum h and 0 otherwise. The term $W_{h.ps}\left(1 - W_{h.ps}\right)$ is the well-known variance of a binomial random variable (Snedecor and Cochran 1971). As before, we can replace the strata weights that are based on sizes of strata in terms of population elements (units) with area-based weights.

The above poststratification estimators are not changed if the initial sample is not obtained under SRS but from a systematic sample. The implicit

proportional allocation is more likely to be satisfied in this case than when the initial sample is obtained under a SRS scheme. According to Cochran (1977, p. 134) poststratified sampling is almost as precise as proportional stratified sampling, provided that the poststratified sample is reasonably large in each stratum ($n_{h.\mathrm{ps}}>20$), and the effects of possible errors in the weights $W_{h.\mathrm{ps}}$ can be safely ignored. As the increase in variance will be small if the average poststratification sample size across strata is sufficiently large, the application of the equations presented here without further adjustments for poststratification is defensible.

3.3.4.6 Pros and Cons of Stratified Sampling

Among the advantages of stratified sampling is the fact that estimates for subpopulation means or totals and their sampling variances are readily available. As the survey procedures for separate strata must be independent (otherwise there may be a covariance between results from different strata) sampling designs and sample sizes can be chosen freely to fit separate strata. In that regard the stratified sampling design is indeed flexible. In almost all cases, a gain in precision of population estimates of means, totals, ratios, and proportions is possible from either a prior stratification or a poststratification of the population.

On the other hand, stratified sampling can also convey certain disadvantages. The effect of inaccurate determination of the sizes of the strata has been mentioned. This problem is manifest when aerial photographs are used as the basis to define strata. Strata boundaries are transferred to the aerial -photographs and the sizes of the strata are determined either planimetrically or through some form of point grid counting (Loetsch and Haller 1964; de Vries 1986). Either way, the procedure is costly and time-consuming and impractical for large-scale surveys. Small spatially scattered strata (leopard pattern) increase the likelihood for errors in the determination of the sizes of the strata. Also, dated aerial photographs make estimates of the sizes of the strata imprecise. In all cases, the effect of the error in the strata area estimation must be carefully considered as it can greatly diminish the gains otherwise expected from stratified random sampling.

While stratification for the estimation of a single population attribute is generally advantageous, the situation is less clear for a multipurpose inventory with its many attributes of interest. For each variable of interest, a different optimum stratification rule is more likely to emerge than not. Consequently, the final stratification becomes a compromise, a compromise flavored by certain threshold and limits of precision not to be imperiled in any attribute or subpopulation.

The utility of stratified sampling is to be particularly critically examined when the inventory is to be repeated on successive occasions. When different sampling designs have been used for the individual strata or the samples have been allocated to strata using a procedure other than proportional allocation, the planning and execution of successive inventories can become very difficult or in extreme cases outright jeopardized. If sample units change from one stratum to another and the inventory design has also changed over time, new strata must be defined to accommodate a possible strata × design effect in the estimates. A resultant dramatic increase in the number of strata becomes a distinct possibility. To the extent it materializes, it affects adversely the design effect of stratified random sampling.

Furthermore, if the population attributes of interest have changed between two inventories the stratification used in the older inventory may be inopportune. An example of such a shift in focus is the current emphasis towards nonproductive functions of the forest at the expense of a narrower focus on timber values.

Most problems that arise from shifting strata and attributes can be mitigated effectively by proportional allocation of sample sizes and a systematic sampling within strata. Inventory designs with these characteristics may be suboptimal for determining an actual condition but they offer the advantage – not to be underestimated – of flexibility and permanence.

3.3.5 Two-Phase Sampling

Two-phase sampling or double sampling is a sampling procedure where two samples are taken from the population. The idea is to exploit an association between the attribute values in the two samples. In the first sample, a large number of easy-to-assess or low-cost sampling units are taken in order to measure one or more auxiliary variables. From this sample a second, smaller, sample is taken for the purpose of assessing the attribute/variable of interest. The statistical link between the auxiliary variable(s) and the variable(s) of interest can be established either by a linear regression (two-phase sampling with regression estimators) or by using the auxiliary variable to estimate the size of the strata (two-phase sampling for stratification). The two-phase design extends naturally to three and more phases (Magnussen 2003).

3.3.5.1 Two-Phase Sampling with Regression Estimators

Two-phase sampling with regression estimators is similar to single-phase sampling with regression estimators, with the difference being that the auxiliary variable, say x, is not measured on all N population elements but on a subsample of

N. In the first phase a large sample of size n_1 is selected; in the second phase a random subsample of size $n_2 \leq n_1$ is selected where the auxiliary variable, *x*, and the variable of interest, say *y*, are measured. The two-phase sample with regression estimator of the population mean for a single variable of interest and a single auxiliary variable is

$$\hat{\bar{Y}}_{2\text{prgr}} = \hat{\bar{Y}}_2 + \hat{\beta} \times \left(\hat{\bar{X}}_1 - \hat{\bar{X}}_2 \right),$$

where $\hat{\bar{X}}_1$ and $\hat{\bar{X}}_2$ are the sample-based estimates of the mean of *x* in the first-phase sample and in the second-phase sample, respectively, $\hat{\bar{Y}}_2$ is the estimate of the population mean of *y* obtained from the second-phase sample, and $\hat{\beta}$ is the least-squares regression coefficient of *y* on *x* computed from the second-phase sample. Note, $\hat{\beta}$ can be improved by recognizing that the first-phase sample of *x* gives a better estimate of the variance of *x* than does the smaller, second-phase sample (Särndal et al. 1992). The second term in the two-phase sampling with regression estimator of the population mean is a term that corrects the SRS estimate by an amount that is proportional to the difference between the first-phase and second-phase estimates of the population mean of the auxiliary variable and the average effect of a one unit change in the auxiliary variable on the expected value of the variable of interest. Unequal probability sampling in the second phase must also be taken into account when the regression coefficient is computed. This is done by weighting the second-phase sample pairs of *x* and *y* by the inverse of their inclusion probability.

An estimator of the variance of $\hat{\bar{Y}}_{2\text{prgr}}$ is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{prgr}} \right) = \hat{\text{var}}\left(y \middle| x \right) \times \left(\frac{1}{n_2} + \frac{\left(\hat{\bar{X}}_2 - \hat{\bar{X}}_1 \right)^2}{\sum_{j=1}^{n_2} \left(x_j - \hat{\bar{X}}_2 \right)^2} \right) + \frac{\hat{\text{var}}\left(y \right) - \hat{\text{var}}\left(y \middle| x \right)}{n_1} - \frac{\hat{\text{var}}\left(y \right)}{N}$$

where $\hat{\text{var}}(y|x)$ is the conditional variance of *y* given *x*, or simply the variance of the expected values of *y* once *x* is known. For a linear relationship $y = \alpha + \beta x + \text{error}$, the conditional variance of *y* is the variance of the linear prediction $\alpha + \beta x$. The first term on the right-hand side of this variance expression contains the variance of the linear predictions of *y*, while the second term adds the variance of the prediction errors. The third term is a correction factor for finite-population predictions. In infinite populations the last term drops out.

An alternative approximation of the variance of $\hat{\bar{Y}}_{2\text{prgr}}$ is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{prgr}} \right) \simeq \frac{\hat{\text{var}}\left(y \right)\left(1 - \hat{\rho}_{xy}^2 \right)}{n_2} + \frac{\hat{\rho}_{xy}^2 \times \hat{\text{var}}\left(\text{y} \right)}{n_1} - \frac{\hat{\text{var}}\left(\text{y} \right)}{N},$$

where $\hat{\rho}_{xy}$ is the second-phase sample-based estimate of the product moment correlation coefficient between the second-phase sample values of y_i and x_i. Again the computational details of $\hat{\rho}_{xy}$ depend on the second-phase sampling design.

Two-phase sampling with regression is typically used in forest inventories where remote-sensing data and field assessments are to be combined While two-phase sampling with regression estimators is often more efficient than two-phase sampling for stratification, it has specific problems when used in practical settings. The cost relationship between the assessment in the first and the second phase is one factor to consider carefully. The other is the strength of the relationship between the variable of interest (y) and the auxiliary variable (x). This strength is often measured in the fraction of the variance in y that is explained by x. The coefficient of determination (ρ^2_{xy}) quantifies this strength; ρ^2_{xy} is a real number between 0 and 1. A higher ρ^2_{xy} means a stronger relationship and conversely a lower variance of the two-phase sampling with regression estimator. In large areas or in forests with a large spatial variability, ρ^2_{xy} values around 0.4 are not uncommon. Thus, only 40% of the variation in y can be explained through variation in x. In homogenous or small-scale forest areas higher R^2 values are commonplace. To expect ρ^2_{xy} values larger than 0.9, however, is unrealistic given the natural variation of the variables and the lack of perfect relationships between common inventory attributes. Furthermore, such high values are probably questionable and could be the result of transformations of x, y, or both or the result of forcing the regression through the origin when an intercept term is significant. The interpretation of estimates of ρ^2_{xy} should always be prudent. A few nontypical observations in the secondary sample or simply a nonrepresentative sample can grossly inflate the sample-based estimates of the population value of the correlation (Royall 2001). A cursory glance at the equation for the approximation of the variance may give the misguided impression that any correlation can be exploited advantageously in two-phase sampling with regression estimators.

De Vries (1986, pp. 117–120) provides an illustrative numerical example of two-phase sampling with regression estimators. In phase 1 the volume in 90 randomly selected photographed plots is determined by a trained interpreter. A subsample of 30 plots is randomly selected and their volume is determined from field measurements of height, diameter, and local volume tables. The correlation between the two volume estimates was strong (0.94) and the 90 photographic interpretations of voume brought about a 60% reduction in the estimated variance relative to a SRS with 30 ground plots.

Attributes measurable in remotely sensed images like tree height, crown diameter, or the number of trees within a defined area are often used as independent variables in a regression function to estimate the growing stock of a forest or forest stand. These types of applied regression estimators, however,

have the distinct disadvantage that the independent variables can only be determined correctly from the aerial photograph under a set of specific conditions and appropriate resolution, conditions that are difficult to meet. Schade (1980), for instance, believes that a scale of 1:10,000 is too small to determine the crown size. In dense multilayered forests, typical in many tropical regions, the assessment of the number of trees or the crown diameter is difficult and in fully stocked stands the direct measurement of tree heights is impossible. Consequently, a double-sampling design for estimation of volume or biomass derived from volume functions, conversion factors, and presumed auxiliary measurements in aerial photographs remains, for many practical applications, a risky proposition.

We have so far regarded the application of two-phase sampling with regression estimators to a single attribute/variable of interest. Forest resource inventories are usually planned and implemented for the estimation of a large number of population parameters, all of interest to owners, managers, and stakeholders. In addition, estimates are often desired for the both the forest as a population and for one or more subpopulations. Detailed representation of estimated parameters might include, for example, a total broken down by tree development stage, by tree species, and by ecological zonation. For each estimate, a new regression relationship has to be derived from the sample data. The independent estimation of a large number of regression models from a single data set invariably entails nonadditive results. Even the probability of obtaining nonsensical results is not trivial. Users of inventory expect and should expect additivity of results. The nonadditivity problem is akin to that encountered in sampling with partial replacement (SPR). The need to derive a multitude of regression relationships and consequently the need to adjust the results to ensure additivity means that the analysis of inventory results based on double sampling with regression estimators quickly becomes very complex if not awkward. Furthermore, the correctness of the regression analysis depends on satisfying a set of rather strong assumptions regarding the residuals and the variables x and y. Also, not all target variables can be estimated with a single estimator. A number of variables on nominal or ordinal scale require a transformation, if possible, to meet the assumptions of the regression model. A nonlinear relationship between x and y would exclude the use of two-phase sampling with regression estimators.

An implicit requirement for the application of regression analysis is that the assessment of the variable of interest and the auxiliary variable is done on the same element. This can only be safeguarded if the sample plots in the two phases coincide exactly. Studies of the positional accuracy in the Swiss National Forest Inventory (NFI) found that the centers of the aerially photographed plots and the terrestrial sample plots were, on average, 5 m apart (Keller 2001). Since great care and much expenditure were applied to obtain

an accurate location of the terrestrial plots in the first Swiss NFI, it is reasonable to assume that their accuracy is the best possible under current practical conditions. Future generations of Global Positioning System (GPS) sensors may raise the bar. In inaccessible forests, with difficult terrain and possibly with crews not fully trained in surveying techniques the accuracy will worsen. In tropical forests sample plot centers rarely coincide with centers marked on a photograph or a satellite image. A close relationship between the correct auxiliary variable and the variable of interest should therefore not be expected. The fact that a relationship between two attributes/variables often changes across locations (Gertner 1984; Walters et al. 1991) further strains the notion of a single linear relationship across the entire sample. Finally, measurement errors in the auxiliary variable attenuate the slope (Fuller 1987), and our estimate of the slope is biased.

In conclusion, in forest inventory a double sampling with regression estimators is only a realistic option for the analysis of a few major high-priority attributes.

3.3.5.2 Two-Phase Sampling for Stratification

Two-phase sampling for stratification is similar to stratified sampling except for the fact that the sizes of the strata are not known without error. Strata sizes are estimated by the larger, first-phase sample and the variable/attribute of interest is assessed in the second phase. In two-phase sampling for stratification an auxiliary categorical variable that can take one of H distinct values is sampled in the first phase for the purpose of estimating the proportion of the population elements/units in each category (stratum). The second-phase sample can be a subsample of units sampled from the first phase (dependent) or an independent sample. We shall limit ourselves to the dependent sample with SRS in each phase. De Vries (1986) treats the rare case of independent sampling in the first and second phases. The national forest inventory in the USA, for example, obtains sample-based estimates of the proportion of the land base that is in the forest stratum, possibly forest stratum (i.e., status is uncertain), and nonforest stratum through an intensive sampling and classification of plots located on aerial photographs (Spencer and Czaplewski 1998). A less intensive ground sampling provides the attributes of interest. The stratum of each ground plot is known at the time of ground sampling. Population parameters are then estimated through a combination of estimates obtained for each of the three strata. The uncertainty surrounding estimates of stratum size has to be included in the estimators of sample variance. In two-phase sampling for stratification with H strata and sample sizes n_1 in the first phase and n_2 in the second phase the estimator for the population mean, for example, is as follows (Cochran 1977):

$$\hat{\bar{Y}}_{2\text{pstr}} = \sum_{h=1}^{H} \frac{n_{1h}}{n_1} \times \hat{\bar{Y}}_h = \sum_{h=1}^{H} \hat{W}_h \times \hat{\bar{Y}}_h,$$

where n_{1h} is the first-phase sample size in stratum h (h=1, . . . , H), $n_i = n_{i1} + n_{i2} + \ldots + n_{iH}, i = 1,2$, and $\hat{\bar{Y}}_h$ is the second-phase estimate of the mean of the hth stratum. We see that the population mean is estimated as a weighted sum of the means of strata with weights (w_h) equal to the first-phase estimates of the relative frequencies of units (elements) in each stratum. A sample-based estimator of the variance of the estimated mean is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{pstr}}\right) = \frac{N-1}{N} \sum_{h-1}^{H} \left(\frac{n_{1h}}{n_1 - 1} - \frac{n_{2h} - 1}{N-1} \right) \frac{W_h \, \hat{\text{var}}\left(y_{2h}\right)}{n_{2h}} + \frac{N - n_1}{N\left(n_1 - 1\right)} \sum_{h=1}^{H} W_h \left(\bar{Y}_{2h} - \hat{\bar{Y}}_{2\text{pstr}} \right)^2,$$

where $\hat{\text{var}}\left(y_{2h}\right)$ is a second-phase sample-based estimate of the variance of variable y in the hth stratum. As expected, the variance is the weighted sum of within-stratum and among-strata variances corrected for sample fractions and a finite population size.

For large $n_1$and N the previous variance estimator approximates the variance in stratified sampling, i.e.,

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{pstr}}\right) \times \sum_{h=1}^{H} \frac{W_h \times \hat{\text{var}}\left(y_{2h}\right)}{n_{2h}} \approx \hat{\text{var}}\left(\hat{\bar{Y}}_{\text{str}}\right)$$

This approximation can often be used with impunity instead of the more complex estimator when the first-phase sample consists of a very large number of classified pixels (say over 1,000) taken from a very large remotely sensed image with, say, over 100,000 population pixels (N). This result ought to be intuitive. With a large sample size in the first phase, the variances of estimates of relative sizes of strata become small and can be neglected and the difference between $\hat{\text{var}}\left(\hat{\bar{Y}}_{2\text{pstr}}\right)$ and $\hat{\text{var}}\left(\hat{\bar{Y}}_{\text{str}}\right)$ has no practical importance.

The advantage of double sampling for stratification over stratified sampling is that a laborious assessment of the sizes of the strata is replaced by a quicker and less costly sampling procedure (Sutter 1990). Strata may be defined exclusively for the purpose of estimation and they may not otherwise form any meaningful subdivision of the population. The within-stratum variance, however, must be smaller than the variance in a nonstratified population before there can be a pay-off from the first-phase stratification in the form of a lower sampling variance. In comparison with double sampling with regression estimators, the derivation of regression functions has been eliminated. In many practical situations there is no suitable auxiliary variable that is uniformly and strongly correlated with the variable of interest across the entire population. Two-phase sampling for stratification would be a logical alternative in these

circumstances. This argument extends to the case where several variables of interest are to be determined and where field samples cannot be located with any great precision or linked with a desired accuracy to observations made in the first phase.

3.3.6 Multiphase Sampling

Two-phase inventories can be extended to multiphase sampling procedures through the inclusion of additional assessment levels (Köhl and Kushwaha 1994). As the number of phases increases the number of possible pairs of two phases that each could be used as an estimation of a desired parameter increases exponentially. The optimum use of the sampled information requires that we combine all possible estimators optimally, i.e., weighted with respect to their sampling variances. While the extension of two-phase estimators to multiphase sampling is theoretically straightforward, the fact remains that the thinning of sample sizes down through the hierarchy of phases quickly erodes any tangible gains in precision by going from two to three or even more phases (Magnussen 2003).

Three-phase sampling with regression estimators is, as expected, a complex design with few practical applications (Pfeffermann et al. 1998). It can hardly be recommended for other than special purpose surveys owing to considerable estimation problems. Therefore, only three-phase sampling for stratification is detailed. As before, we limit details to random subsampling in the second and third phases Examples of practical applications of three-phase sampling for stratification are given in Cherrill and Fuller (1994), Kirkman (1996), Williams (1996), Lunetta et al. (1998), Vogelmann and Howard (1998), Cannell et al. (1999), Lunetta and Elvidge (1999), Brown et al. (2000), Cruickshank et al. (2000), Flores and Martínez (2000), Moran et al. (2000), Chong et al. (2001), Franklin (2001), and Magnussen (2003). As general references Bowden (1979) and Johnston (1982) can be recommended.

Estimators for three-phase sampling for stratification use an extension of the notation of two-phase estimators. The first two phases provide information for estimation of the relative frequencies of the strata in the population. This includes estimation of the proportions of second-phase strata in each of the first-phase strata. Given N, these estimators of strata proportions can be used to estimate the number of population units in each stratum and the combination of first-phase and second-phase strata. These estimates of the sizes of the strata are needed to scale estimates from subsamples in the second and third phases to estimates for first-phase sampling. In summary, separate estimators of the population parameter of interest are produced for each phase and then combined to give one final estimate. The unbiased estimator of the population

mean for L first-phase strata and H second-phase strata and sample sizes n_1, n_2, and n_3 in the three phases is

$$\hat{\bar{Y}}_{3\text{pstr}} = \sum_{l=1}^{L} \hat{W}_l \sum_{h=1}^{H} \hat{W}_{lh} \times \hat{\bar{Y}}_{3lh} = \sum_{l=1}^{L} \hat{W}_l \times \hat{\bar{Y}}_l,$$

where $\hat{w}_{lh}$ is the proportion of second-phase sampleøs in first-phase stratum l and second-phase stratum h, and $\bar{Y}_{3lh}$ is the third-phase mean of sample values of y in first-phase stratum l and second-phase stratum h. Again, we see that the mean is simply a doubly weighted average of third-phase estimates with weights equal to the relative stratum frequencies in the first and second phases. First-phase sample sizes in first-phase strata l and second-phase strata h are n_{2lh} with $l = 1,\ldots, L$ and $h = 1,\ldots, H$. The total number of second-phase samples in first-phase strata l is $n_{2l} = \sum_{h=1}^{h} n_{2lh}$. Under SRS in each phase, simple ratios of sample sizes are used to estimate the number of population units in each stratum. For example, if we have n_{2lh} second-phase samples out of a total of n_{2l} second-phase samples in first-phase strata l and second-phase strata h, and we have n_{1l} first-phase samples out of a total of n_1 first-phase samples in first-phase strata l, and we would estimate the number of population units in the $l \times h$ strata to be $N \times (n_{1l}/n_1) \times (n_{2lh}/n_{2l})$. After these preliminaries we can write the approximate variance of the estimate of the population mean as

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{3\text{pstr}}\right) \simeq \hat{\text{var}}_{\text{among first-phase strata means}} + \hat{\text{var}}_{\text{among second-phase strata means}} + \hat{\text{var}}_{\text{within first-phase} \times \text{second-phase}}$$

where

$$\hat{\text{var}}_{\text{among first-phase strata means}} = \frac{N(N-1)(n_1-1)}{n_1^2(N-1)^2} \sum_{l=1}^{L} \hat{W}_l \left(\hat{\bar{Y}}_l - \hat{\bar{Y}}_{3pstr}\right)^2,$$

$$\hat{\text{var}}_{\text{among second-phase strata means}} = \frac{N(N-1)(n_1-1)}{n_1^3(N-1)^2} \sum_{l=1}^{L} (n_{1l}-1)\tilde{V}_l,$$

and

$$\hat{\text{var}}_{\text{within first-phase} \times \text{second phase strata}} = \frac{N-n_1}{n_1(N-1)} \sum_{l=1}^{L} \hat{W}_h \left(\hat{W}_h - \frac{N-n_1}{n_1(N-1)}\right) \times \ddot{V}_l,$$

with

$$\tilde{V}_l = \left[\ddot{V}_l - \frac{1}{n_{2l}} \sum_{h=1}^{H} \hat{W}_{lh} \times \left(\frac{n_{2l}}{n_{2lh}} - 1\right) \Big/ \left(\frac{1}{n_{2lh}} - \frac{1}{n_{1l}}\right) \times \hat{\text{var}}\left(y_{hl}\right)\right],$$

and (Cochran 1977)

$$\tilde{V}_l = \frac{n_{1l}-1}{n_1} \sum_{h=1}^{H} \left[\frac{N \times \frac{n_{1l}}{n_1} \times \frac{n_{2lh}}{n_{2l}} - 1}{N \times \frac{n_{1l}}{n_1} - 1} - \frac{n_{2lh}-1}{n_{2l}-1} - \right] \frac{\hat{W}_{lh}}{n_{2lh}} \hat{var}\left(y_{lh}\right) + \ddot{V}_l,$$

and

$$\ddot{V}_l = \frac{N \times \frac{n_{1l}}{n_1} - n_{1l}}{n_{1l}\left(N \times \frac{n_{1l}}{n_1} - 1\right)} \sum_{h=1}^{H} \hat{W}_{lh}\left(\hat{\bar{Y}} - \hat{\bar{Y}}_l\right)^2.$$

3.4 Errors in Forest Surveys

In sampling surveys, two types of errors are distinguished: sampling errors and nonsampling errors. Sampling errors result from the facts that only part of the population is surveyed and population parameters are estimated from the sample. The estimated parameters may deviate from the true population values. One way in which sampling error may be expressed is through the standard error of the mean. This ought to be given for all estimators, as it is essential for the correct interpretation of inventory data. Nonsampling errors, on the other hand, are not connected with the problem of dealing with only part of the population but arise from inaccurate measurements, less-than-perfect measuring devices, mistakes in the execution of the sampling plan or protocol, and sampling the wrong units/elements. Nonsampling errors of this nature are likely to inflate the apparent sampling variance and to introduce a bias in the estimates. When the sample observations are derived from model-based predictions and not a direct measurement per se our data are subject to model error. Since models only predict the expected value, the apparent sampling variance of model-based predictions will be too low; the residual model variance has been left out.

From a statistical point of view the reliability of results can be quantified by giving their precision, accuracy, Mean Square Error (MSE), or bias. We shall give definitions as there continues to be considerable confusion about the correct use of these terms:

Precision: Precision refers to the size of deviations in the estimate of a population parameter in repeat application of a sampling procedure. The standard error or confidence interval quantifies precision. Increasing the number of observations increases the precision of a statistical estimate.

Accuracy:Accuracy refers to the size of deviations between an observed value and the true value. Thus, if we know the true value of a population parameter then we can also define the accuracy of a survey estimate as the deviation between the estimate and the true value.

Bias: Bias is directly related to the accuracy of an estimate. Bias is the difference between the estimated value and true value of a parameter. We cannot estimate bias unless we know the true value of a parameter. In practice we do not have this knowledge.

The effect of precision and bias can be seen in Fig. 3.11.

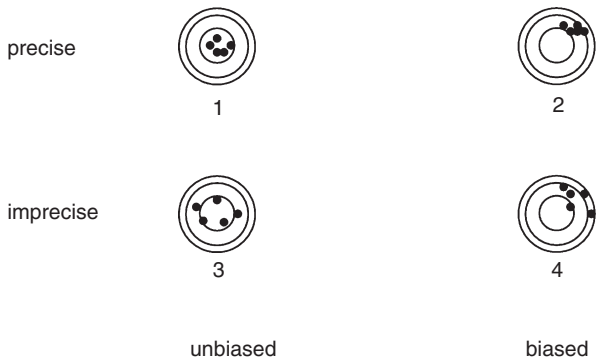


Fig. 3.11. Example for the concept of accuracy and precision of an estimator. If we assume that the target center is the true but unknown population value and that the position of each shot represents the estimate obtained by a random sample, it follows that (*1*) the estimator is precise and accurate, (*2*) the estimator is precise but biased, (**3**) the estimator is imprecise but unbiased, (*4*) the estimator is imprecise and biased. (After Vanclay 1994)

Mean Square: MSE is a useful measure of reliability. It combines the precision of an estimate with the square of the bias. The MSE is a useful criterion for the comparison of two or more competing estimators, possibly with different amounts of bias. A direct comparison of estimators in terms of precision only may skew the choice towards estimators that generate highly precise but biased estimates. According to Cochran (1977, p. 15) the MSE of an estimate of, say, a population total is

$$\mathrm{MSE}\left(\hat{Y}\right) = \hat{v}ar\left(\hat{Y}\right) + \left(\hat{Y} - Y\right)^2.$$

For unbiased estimators MSE and precision are asymptotically $\left(\{n, N\} \to \infty\right)$ identical. In the following chapters mostly unbiased estimators for population parameters will be presented. The MSE and the precision of the estimates derived from unbiased estimators should, therefore, be asymptotically equivalent. Alternative estimators such as, for example, Bayesian and Stein estimators, seek an attractive balance between bias and precision, often pursuing a minimum variance at the expense of some small amount of bias (Box and Tiao 1973; Congdon 2001).

The standard procedure for calculating sampling error does not allow for the influence of bias. Nevertheless, bias may multiply the sampling error by several magnitudes (Gertner 1984). Increasing the sample size may certainly decrease the sampling error but it can also increase the relative influence of bias. Consequently, possible bias should be guarded against from the earliest stage of planning a survey. It is often possible to somehow assess various sources of inventory errors (Gernter und Köhl 1992) and then gauge the effect

of potential bias arising from different sources during the inventory or perhaps owing to the choice of estimators. Simulation studies may be needed to quantify bias introduced by an estimator.

3.4.1 Non-Sampling inventory errors

Estimators of population attributes and their sampling variances have so far been presented as if the observations were not only complete but also the best possible. Best possible means that the most accurate technique for obtaining data is applied everywhere. Practical surveys can rarely live up to this ideal; forest inventories are no exception (Goelz and Burk 1996; Chen 1998; Lesser and Kalsbeek 1999; Chen and Cowling 2001).

3.4.1.1 Nonobservation

The sample can be incomplete owing to nonobservation of sample units or errors in the population frame from which sample units are selected (Särndal et al. 1992). Nonobservations of sample units can occur when some units are not visited because (1) access is denied, (2) access poses a danger to the survey crew, (3) the sample unit could not be located, and (4) sampling was terminated owing to time and budgetary constraints. Errors in the population frame usually result in an undercoverage, certain population units have a zero probability of inclusion in the sample, or conversely an inflated inclusion probability if the unit appears more than once in the frame. Regardless of cause, an incomplete sample will, as a rule, result in a biased estimate of both the population attribute value and its sampling variance. We can easily appreciate this result if we subdivide a population of N units into two parts, one with attribute values y_{1i}, $i = 1,\ldots, N_1$ for which observations can or will be made, and the second with attribute values $y_{2i}, i = 1, ..., N_2$ for which observations will be missing regardless $(N_1 + N_2 = N)$. Let us assume $w_2 = N_2 \times N^{-1}$, our average sample-based estimator of, say, a population mean is biased since $E_s(\hat{\bar{Y}}) - \bar{Y} = w_2(\bar{Y}_1 - \bar{Y}_2)$, where the expectation is with respect to the sampling distribution of sample estimates. Only in the rare case when observations are missing completely at random (MCAR; Little and Rubin 1987) can we expect no bias since under MCAR Pr (Y | missing) = Pr (Y) × Pr (missing) and $E(\bar{Y}_1) = E(\bar{Y}_2)$. To mitigate a potential bias from missing observations we must either make assumptions about either y_2 or the mechanism leading to a missing observations, or perhaps both. We can impute the missing values by suitable substitutes by specifying a data model for y_2 to complete the sample and then obtain the usual estimates as if the sample was complete. There are many ways to do the imputation and to obtain the statistical estimators from samples with imputations, each relying on

a specific set of assumptions. Readers are referred to Rubin (1987), Efron (1994), Schafer and Scheinker (2000), and McRoberts (2001) for details. Alternatively, if data are missing at random and there exists a quantifiable relationship between data values and the probability of a missing value we may be able to adjust our sample estimates by obtaining a new set of sample inclusion weights from estimates of the probability of a missing value $\Pr(Y_1, X_1, X_2)$, where x is an auxiliary variable known for all population elements. The adjustment option is only available if the sample contains at least some elements from across the entire range of x. For example, if the probability of obtaining a sample is a monotone function g of the slope ϕ of the terrain at the sample location and $0 < g(\phi) < 1$ then the inverse $1/g(\phi)$ can be used in conjunction with the original design-based inclusion probabilities to reduce the bias due to missing observations.

Estimators of sampling variance obtained from surveys with missing data do not account for a potential bias. The MSE (MSE = var + bias^2) would be a preferred estimate of precision in the presence of a potential bias; however, we will rarely know the magnitude of the bias. A rule of thumb (Cochran 1977) says that unless the absolute bias is less than 0.1 times the sample-based estimator of the standard error of an estimate, the reported standard error and any conventional confidence interval for the estimator could be seriously misleading.

In sampling for proportions we can at least impose limits on the missing data and use these limits to construct a conservative confidence interval for the estimated population proportion. Cochran (1977) gives an example with $n_2 = 200$ missing binary observations out of $n = 1{,}000$ target observations (i.e., $n_1 = 800$) and 80 positive responses ($y = 1$). By assuming that the missing data at one extreme would be all 0 and all 1 at the other extreme, the following conservative 95% confidence interval of the sample estimate of the population proportion becomes

$$Cl\left(\hat{P} = 0.08\right) = \left\{\frac{80 + 0}{1000} - 1.96 \times \sqrt{\frac{0.08 \times 0.092}{1000}}, \frac{80 + 200}{1000} + 1.96 \times \sqrt{\frac{0.08 \times 0.092}{1000}}\right\}$$
$$\cong \{0.063, 0.308\}$$

We can apply the same rationale towards estimating the sample size needed to achieve a given target precision when it is known that some binary observations will be missing. Since bias does not decrease with sample size, even a modest rate of nonresponse (less than 10%) can have a serious impact on the quality of survey estimates and every possible effort should be made to obtain a complete sample.

3.4.1.2 Measurement Errors

Directly observed or compiled attribute values of a sampled population element (unit) are rarely, if ever, completely free of errors or bias. An observation deviating from the best possible observation is said to be in error and possibly biased. Conceptually we can write an observed, viz., a compiled, attribute value

y'_i for the ith population element as a linear sum of the best possible value y_i plus a series of error and bias terms arising from various sources (k). For a univariate attribute we can write our observations (compilations) as

$$y'_i = y_i + \sum_k e_{ik} + \sum_k b_{ik} = y_i + e_{i\cdot} + b_{i\cdot}, E(e_{ik}) = 0, b_{ik} \neq 0,$$

where $e_{i\cdot}$ *and* $b_{i\cdot}$ denote the sum of error and bias in the observation (compilation) y'_i. The errors e_{ik} will depend on the attribute and on the entire process, including the design that generates the observed, viz., compiled, values y'_i. Extension to multivariate attributes is straightforward. Only the univariate case will be detailed.

For surveys with errors and possibly bias in observed/compiled attribute values the sample-based estimate of, say, a population mean $\bar{Y}$ is $\hat{\bar{Y}}' = \hat{\bar{Y}} + \bar{e} + \bar{b}$, i.e., the estimate we would have obtained if there were no errors or no bias $\left(\hat{\bar{Y}}\right)$ plus the sum of the average error and the average bias. We would normally not know the average error nor the average bias, but it is often assumed that the errors are independent and cancel across the sample, which would leave us with an estimate with a bias of $\bar{b}$. With these assumptions the expected MSE of a sample estimate is

$$E_s\left[\text{MSE}\left(\hat{\bar{Y}}'\right)\right] = \text{var}\left(\hat{\bar{Y}}\right) + \text{var}\left(\bar{e}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(b_{i\cdot} - \bar{b}\right)^2 + 2 \times \text{cov}\left(\bar{e}, \hat{\bar{Y}}\right),$$

where expectation is over the sampling distribution of $\hat{\bar{Y}}'$. The expected MSE is thus the sampling variance in the absence of errors and bias $\left[\hat{\text{v}}\text{ar}\left(\hat{\bar{Y}}\right)\right]$ plus the sum of the variance of the average error, the average of the squared bias differential of individual samples, and twice the covariance between the average error and $\hat{\bar{Y}}$. The covariance term accounts for a possible correlation of errors and attribute values. For example, a given surveyor may introduce a constant observer bias to all samples assigned to this individual. Likewise, an instrument may introduce an error that is a function of the read-out value.

The practical consequence of the expected MSE equation is that only the first two terms decline at a rate of $1/n$; the bias contribution is usually independent of n but may actually increase with n if an increase in sample size somehow compromises the quality of measurements or compilations. The covariance term may decline with increasing n but only under a set of rather specific and restrictive circumstances. Consequently, even a modest bias or error covariance can greatly inflate the relative standard error of a survey. Gertner and Köhl (1992) gave a clear demonstration of this in their analysis of the errors and possible bias in the Swiss NFI. A modest 1% bias in the volume compilations, for example, would double the relative standard error of the estimate of the total volume in a strata dominated by Norway spruce. When the covariance term $\text{cov}\left(\bar{e}, \hat{\bar{Y}}\right)$ is negative, the observed variance can be less than the actual sampling variance of the best possible estimate. When the

observable values are constrained or categorical, the covariance term is generally negative.

Unless we somehow produce estimates of the measurement error variance, the bias, and the error covariance, we will not know by how much our sample-based estimates of population attributes and estimates of sampling error have been biased. If we are willing to assume a constant bias and no error covariance the variance of $\hat{\overline{Y}}'$ is

$$\mathrm{var}\left(\hat{\overline{Y}}'\right) = \frac{1}{n}\,\hat{\mathrm{v}}\mathrm{ar}\left(\hat{\overline{Y}}\right)\left(1 - \frac{n}{N}\right) + \frac{1}{n}\,\hat{\mathrm{v}}\mathrm{ar}\left(e\right).$$

Thus, the expected sampling variance of a real-valued population attribute is inflated by the variance of measurement errors; the bias term vanishes since it was assumed to be constant. On the other hand, the expected sample-based estimate of the sampling variance is

$$E\left[\hat{V}ar\left(\hat{\overline{Y}}'\right)\right] = \left[\frac{1}{n}\,\hat{V}ar\left(\hat{\overline{Y}}\right) + \frac{1}{n}\,\hat{V}ar\left(e\right)\right]\left(1 - \frac{n}{N}\right)$$

Hence, a sample-based estimate of the sampling variance of a population attribute observed/compiled with an error and possibly a constant bias is only approximately unbiased if the sample fraction n/N is negligible.

Measurement errors may, however, not be independent across samples. In presence of a correlation of some or all measurement errors (and constant bias) the expected sample-based estimate of sampling variance becomes

$$E_s\left[\hat{\mathrm{v}}\mathrm{ar}\left(\hat{\overline{Y}}'\right)\right] = \left[\frac{1}{n}\,\hat{\mathrm{v}}\mathrm{ar}\left(\hat{\overline{Y}}\right) + \frac{1}{n}\,\hat{\mathrm{v}}\mathrm{ar}\left(e\right)\left(1 - \hat{\rho}_e\right)\right]\left(1 - \frac{n}{N}\right),$$

where the expectation is over the sampling distribution and $\hat{\rho}_e$ is an estimate of the intrasample correlation between errors of measurement with

$$\hat{\rho}_e = E_s\left[\left(e_i \times e_j \middle| \{i,j\} \in s\right)\right] \times \hat{\mathrm{v}}\mathrm{ar}\left(e\right)^{-1}.$$

Since the correlation is mostly positive a sample-based estimate of sampling variance appears more precise than it is. Positively correlated errors are not unusual in attributes derived from remotely sensed data (Since we have ignored the covariance of errors in our calculations of variance) (Congalton 1988; Dobbertin and Biging 1996).

3.4.1.3 Estimating Nonsampling Errors and Bias

Bias and measurement errors can seriously compromise the quality and precision of a survey. Diligence and high standards are required in all aspects and all phases of a forest inventory to keep measurement errors and bias within acceptable bounds. Quality standards and quality checks are integral parts of a forest inventory. Still, measurement errors and bias are virtually impossible to

stamp out of a forest inventory. It is therefore good practice to investigate the impact of measurement errors and bias on survey results. An error budget discloses all possible sources and the expected impact of error in the entire process that begins with a visit to a sample unit and ends with a set of survey estimates of population attributes and is ideally suited for this purpose (Gertner and Köhl 1992; Kangas 1996). The error budget will ideally not only disclose the possible bias in estimators but also suggest where and how better standards can mitigate the impact of bias and measurement errors. In well-designed inventories with high measurement and compilation standards the contribution of natural intrinsic variation in attribute values to the overall sampling variance is usually orders of magnitude larger than the contribution of measurement errors and bias (Gertner and Köhl 1992; Kangas 1996).

Survey observations (viz., compilations) include measurement errors and bias. Estimates of the measurement errors and possible bias can be obtained either by repeated measurements and compilations performed on a subset of sample units or by model-based Monte Carlo simulations of the entire process that produced the desired estimates.

The repeated measurement option is simple but costly, and also potentially flawed unless great care is taken to ensure that the repeated measurements allow an unbiased estimation of the error structure and possible bias. Several pairs of two independent repeat observations y'_{i1} and y'_{i2} of the (best possible) sample attribute value y_i in the ith sample unit allow, under the assumption of equal bias and equal error variance, the estimation of the measurement error variance and the covariance of errors associated with a single sample. Estimates are obtained by solving for var(e) and cov (e_1, e_2) the following equations:

$$\frac{\sum_{i=1}^{m}\left(y'_{i1}-y'_{i2}\right)^2}{m}=2\times\left[\operatorname{var}\left(e\right)-\operatorname{cov}\left(e_1,e_2\right)\right]$$

and

$$\frac{\sum_{i=1}^{m}\left(y'_{i1}+y'_{i2}\right)^2}{m}=2\times\left[\operatorname{var}\left(e\right)+\operatorname{cov}\left(e_1,e_2\right)\right],$$

where m is the number of representative sample units selected for repeat observation. Under SRS, n large, and $m = 0.5 \times n$, the average of two replicate estimates of the sampling variance provides an approximate estimate of the sampling variance of $\hat{\bar{Y}}$ via

$$\hat{\operatorname{var}}\left(\hat{\bar{Y}}\right)\cong 0.5\left[\hat{\operatorname{var}}\left(\hat{\bar{Y}}'_1\right)+\hat{\operatorname{var}}\left(\hat{\bar{Y}}'_2\right)\right]-\hat{\operatorname{var}}\left(e\right)\left(1-\hat{\rho}_e\right).$$

Fewer than 30 repeat samples will not provide reliable estimates (Cochran 1977). The potential of a bias in the estimate of the measurement error variance arises from the fact that if the same person is asked to do a task twice a recall from the first execution is likely to influence the second execution in

some way or other. If a different person is assigned to the second observation the estimate of the measurement error variance will be confounded by interpersonal bias. A direct sample-based estimate of bias can only be obtained from multiple repeat observations (more than six) of an attribute value using the adopted survey methodology and one final observation of the best possible value. The large number of repeat measurements ensures that the average measurement error will be close to zero and therefore $\hat{b}_i \cong \overline{Y'_i} - y_i$.

In forest inventory the repeat measurement approach to estimate measurement error variance and bias is often either impractical or too costly. Instead the surveyor attempts to produce model-based estimates of measurement errors based on Monte Carlo simulations. The measurement error budget starts with the estimator of the (best possible) attribute value of interest. The estimator is then expanded into a model that includes variables for all the basic attribute values observed, viz., measured, on a sample unit. The model also includes all functions and their parameters needed to transform basic sample attribute values into the desired attribute value. Once the model has been established, a measurement error distribution and possible bias are specified for each variable and parameter in the model based on either results from specialized studies or subject knowledge. The actual survey data are then often assumed to be the best possible (no measurement error and no bias). Errors and bias are then added to all sample observations and model parameters according to the specified models in order to simulate a new data set from which the desired estimates are obtained. By repeating this process a large number of times (more than 500) one can estimate the relative contribution of bias and measurement errors to the overall estimate of sampling error since we "know" both the best possible value and the simulated observed value of all inputs to an estimator. An example of a measurement error budget for volume estimation follows since it will contain most of the commonly encountered features and problems.

Let us assume that we are interested in estimating the total volume of stem wood in a population area (PA) of 300-ha (strata) from a simple random sample of size $n = 40$ with fixed-area circular sample plots with a nominal area of $A = 100$ m^2 for all plots. In each plot we measure the diameter at breast height (DBH) of the n_{ti} trees in the ith plot for which DBH≥5 cm with a tape. The height of a maximum of 12 trees representing the range of DBH in the plot is measured on each plot. A function that predicts tree height (HT) from DBH is estimated from pairwise observations of HT and DBH. A predicted height $\widehat{HT}$ is obtained for all trees with no measured height. A volume equation transforms paired values of DBH and HT, viz., DBH and $\widehat{HT}$, into a stem-volume prediction of a single-plot tree. Stem volumes of single trees are summed on a per plot basis and expanded to plot-specific estimates of the population total. The desired estimate is the average plot-specific estimate of the total and the sampling variance of the estimated total is the variance of the plot-specific

estimates of the total divided by n and multiplied by the finite-population correction factor. With no measurement errors and no bias the ideal (theoretical) sample estimate of total stem volume with DBH≥5 cm would be

$$\widehat{\text{vol}} = \frac{\text{PA} \times 10^4}{A} \sum_{i=1}^{n} \sum_{j}^{n_{ti}} \hat{\alpha}_{oij} + \hat{\alpha}_{1ij} \text{DBH}_{ij}^2 + \hat{\alpha}_{2ij} \text{HT}_{ij} + \hat{\alpha}_{3ij} \text{DBH}_{ij}^2 \text{HT}_{ij},$$

where vol is the total volume of stem wood and α_{kij}, $k = 0,...,3$ are the best possible parameters of a volume equation that generates the best possible predictions of stem volume for the jth tree in the ith plot. All design variables (PA, A, n), and all variables and parameters used in the volume compilation $\{n_{ti}$, DBH, HT, $b_{0ij}, \ldots b_{3ij}\}$ are subject to both measurement errors and bias. The actual observed volume estimate is based on

$$\widehat{\text{vol}'} = \text{PA}' \times 10^4 \sum_{i=1}^{n} \frac{1}{A'_i} \sum_{j}^{n_{ti}} \hat{\alpha}'_0 + \hat{\alpha}'_1 \text{DBH}'^2_{ij} + \hat{\alpha}'_2 \text{HT}'_{ij} + \hat{\alpha}'_3 \text{DBH}'^2_{ij} \text{HT}'_{ij},$$

where, as befor e, X' is an attribute value with measurement error and possible bias and X is the best possible counterpart with no error and no bias. Note that the PA may also be in error and that only one volume equation is used to predict stem volume from DBH and HT. We could have used plot-specific volume equations or volume equations specific to various subsets of plots (Lappi 1991) but to keep this example relatively easy we opted for just a single equation. The error-free and bias-free volume estimate and the actual observed estimate are linked through the following set of measurement equations (carets have been suppressed to avoid cluttering):

$$\begin{aligned}
&\text{PA}' = \text{PA} + e_{\text{PA}},\\
&A'_i = A + e_i^2(r) \times \pi + 2e_i(r)\sqrt{\pi},\\
&n'_{ti} = n_{ti} + e_i(n_{ti}),\\
&\alpha'_0 = \alpha_0 + e_i(\alpha_0) + e_{ij}(\alpha_0) + b_i(\alpha_0) + b_{ij}(\alpha_0),\\
&\alpha'_1 = \alpha_1 + e_i(\alpha_1) + e_{ij}(\alpha_1) + b_i(\alpha_1) + b_{ij}(\alpha_1),\\
&\alpha'_2 = \alpha_2 + e_i(\alpha_2) + e_{ij}(\alpha_2) + b_i(\alpha_2) + b_{ij}(\alpha_2),\\
&\alpha'_3 = \alpha_3 + e_i(\alpha_3) + e_{ij}(\alpha_3) + b_i(\alpha_3) + b_{ij}(\alpha_3),\\
&\text{DBH}'_{ij} = \text{DBH}'_{ij} + e_i(\text{DBH}) + e_{ij}(\text{DBH}) + b_i(\text{DBH}) + b_{ij}(\text{DBH}),\\
&\text{HT}'_{ij} = \begin{cases} \text{HT}_{ij} + e_i(\text{HT}) + e_{ij}(\text{HT}) + b_i(\text{HT}) + b_{ij}(\text{HT}) \text{ if measured} \\ \text{HT}_{ij} + e_i(\widehat{\text{HT}}) + e_{ij}(\widehat{\text{HT}}) + b_i(\widehat{\text{HT}}) + b_{ij}(\widehat{\text{HT}}) \text{ if predicted} \end{cases},
\end{aligned}$$

where $e(X)$ denotes a measurement error of attribute or parameter X with an expected value of zero and $b(X)$ the bias of attribute X. Subscripts refer to plot level (i) and tree level (j). No covariance between any of the errors or error processes was specified in the equations. They were left out intentionally. One should, however, expect covariance between some errors and between some

error processes but they would be inventory-specific and we wished to keep the example simple. A model-based transformation of basic observations invariably introduces covariance between errors and processes. Kangas (1996) illustrates how one can use Taylor series approximations to obtain model-consistent correlated multivariate errors for the Monte Carlo simulation. Another approach to obtain correlated errors is to specify only a set of basic errors (as in n_{ti}, DBH, and HT) and then propagate these errors through the estimation process using, again, Taylor series approximations to the best possible estimate or conversely the observed estimate. Dedicated textbooks give the necessary details (Fuller 1987; Goodchild and Gopal 1989; Carroll et al. 1995).

To arrive at an error budget we need to specify the distribution of all measurement errors at both the plot level and the tree level and all bias terms and then conduct Monte Carlo simulations of repeat sampling of actual+error+bias attribute values followed by an estimation of total volume for each repeat sample. It is customary to perform a sequence of Monte Carlo simulations, each with a specific set of errors and biases set to zero. This allows a separate assessment of various sources of errors and bias.

Plot-level errors and bias are determined by plot-specific characteristics ranging from topographical attributes to stand/forest conditions, in general, and plot-specific aspects of the measurement process, in general. For example, the plot area may be in error owing to an error in the slope correction to a horizontal reference area (Gertner and Köhl 1992). The number of plot trees may be in error because inappropriate corrections were done to adjust for boundary effects (Gregoire and Scott 2003) or mistakes were made when it was decided whether a tree located at the plot boundary was inside or outside of the plot. Change in surveyors, surveyor diligence, weather conditions, and time of day may also introduce plot-specific errors and/or bias.

3.5 Selection of Trees on Sampling Units

The sampling units of forest inventories are usually not individual trees but rather a group of trees satisfying some criterion. Sample trees may be selected by, for example, satisfying the criterion of location inside a sample plot, exceeding a distance-weighted size threshold in point sampling, proximity to a survey location or a survey line, or satisfying a rank proximity criterion at a sample location.

The rank proximity criterion includes a fixed number of trees closest to a sample location. For example, the six trees closest to a random sample location may be selected (Prodan 1965). Estimators based on this type of tree selection

will often be biased, especially if the spatial distribution of trees is aggregated or in some other way displays distinct spatial patterns. Further, in dense or trackless forests it is time-consuming and expensive to determine the ranking of tree distances to a random point. Consequently, this procedure is not to be recommended for tropical forests and is not discussed here further. Those who would like to know more about it are referred to the literature (Prodan 1965; Payandeh and Ek 1986; Pollard 1971).

3.5.1 Tree Selection with Fixed-Area Sampling Units

Fixed-area sampling units are the simplest intuitive basis for selecting trees to be assessed in forest inventories. The term plot is applied to small circular, rectangular, square, or triangular areas. A strip is a rectangular sample area, whose length is a multiple of its width. Unbiased estimates can be computed for all sample areas, no matter what their shape. In planning an inventory, survey costs must be weighed against desired precision to determine the optimum size and shape of the sample plots.

The shape of a sample plot is mainly determined by cost and other practical considerations. In temperate latitudes, circular plots are usually employed as having the smallest periphery in relation to area and consequently the lowest number of borderline trees. In tropical forests, where the undergrowth hinders both access and visibility and large areas must be surveyed, it is usual to take rectangles or squares because such plots are the easiest to establish. Very often strips of up to 30-m wide and several hundred meters long are recommended.

For a fixed total sample area, choosing a larger plot size means that the sample size goes down since the product of sample size and plot area is constant. A large plot is likely to produce a lower among-plot variance than a smaller plot since large plots in general display more within-plot variation than do smaller plots. Yet the lower number of larger plots afforded under a fixed total sample area may actually produce a higher standard error than sampling with a smaller plot (Correll and Cellier 1987; Magnussen 1998; Gray 2003). The optimum plot size in terms of minimum sampling variance for a fixed total sample area is determined by the spatial distribution and the variability of the forest to be surveyed. Small plots in homogeneous forests may furnish results with higher precision, as the number of independent observations for a given sampling intensity is higher. On the other hand, in heterogeneous forests, the coefficient of variation between small plots may increase so greatly that it would be better to use a larger plot. Consequently, not only the costs but also the variability of the inventory area must be taken into consideration. A key

statistic to gauge the efficiency of different plot sizes is the intraplot correlation coefficient (Cochran 1977). The coefficient measures the similarity of observations within the sample plot. Basically, the more similar the observations are, the more efficient is the small plot, and vice versa (Correll and Cellier 1987; Saborowski and Smelko 1998).

Zeide (1980), as well as Mesavage und Grosenbaugh (1956), Tardif (1965), and O'Regan and Arvantis (1966), examined various methods for optimizing plot size. Zeide weighed the time needed to locate a plot against the specified precision and stated that the optimum plot design is the design with the lowest expenditure for the specified precision. The optimum plot size a_{opt} was computed from

$$a_{\text{opt}} = a_1 \left(\frac{t}{m} \right)^2 ,$$

where a_1 is the size of the plot used in a pilot survey, t is the average travel time between two neighboring plots, and m is the average measuring time on a plot of size a_1. Zeide concluded from this that the greater the distance between plots, the larger the plots should be.

To compare two plot designs with plot types 1 and 2, the relative efficiency of type 1 for the estimation of, say, a population total is

$$\text{eff}_{\text{type 1: type 2}} = \frac{\hat{\text{var}}\left(\hat{Y} \middle| \text{plot type 1} \right) \times \text{time of inventory with type 1}}{\hat{\text{var}}\left(\hat{Y} \middle| \text{plot type 2} \right) \times \text{time of inventory with type 2}} .$$

An efficiency ratio less than 1 means that plot type 1 is more efficient, and vice versa. Note that the efficiency depends on the population parameter of interest. It is possible that one plot type is more efficient for one parameter but less efficient for a different parameter.

As a rule of thumb, a plot should be large enough to contain enough trees per plot for the survey results to be representative of the population at large while at the same time keeping the time to complete a plot to a minimum. It follows that small plots should be employed for dense stands with small trees, and large ones for open stands and large trees. Very often, a distinction is made between unproductive relocation time between plots and productive survey times. When travel time is significant, as in a tropical forest, the size of inventory plots tends to be large, often in the 0.4–0.5-ha range.

The horizontal plane is the reference base for all inventory data, and sample plots in sloping terrain must be adapted accordingly. Three distinct procedures for this adaptation are given next; the first is general, while the second and third are for circular plots only:

1. All plots: Demarcate the plot on the incline and then expand the plot proportionally to the degree of inclination in such a way that an orthogonal projection of the expanded plot onto the horizontal plane matches exactly the nominal plot in area.
2. Circular plots: Demarcate an ellipse on a slope with the short axis and the long axis determined in such a way that when the ellipse is projected orthogonally onto the horizontal plane it coincides with the circular outline of the nominal inventory plot.
3. Circular plots: Measure or compute the horizontal distance of candidate plot trees to the plot center. Only trees within a distance of the nominal plot radius are included and measured.

In order to simplify the field survey and to facilitate an audit and a quality control of the inventory the enlargement of the plot in proportion to the degree of inclination appears most suitable. Köhl and Brassel (2001) showed for the Swiss NFI that there is no significant difference between the second and third expansion approach. In mountainous regions, an error in the corrected plot can induce a nontrivial error in the survey results. For this reason, corrections for slope inclination must be made by fully qualified personnel only, and all pertinent data about the expansion/correction process should be captured to allow control and possibly correction of the errors.

Sample plots in areas with a high stem density may contain a large number of trees. The aforementioned general principle about small plots being more efficient in areas with a high tree density and larger plots being more efficient in areas with a low tree density has led to a design with multiple concentric plots. Two or more plots of differing size are demarcated around a given sample point. In the smallest area, all trees with a diameter greater than a given minimum fixed by design (e.g., 12 cm) are surveyed. In the larger plots, the minimum diameter threshold is higher. This design often allows a considerable reduction of survey time with barely noticeable decreases in efficiency. Figure 3.12 shows the concentric plot design employed in the Swiss NFI. On the smaller, 200-m^2 plot, all trees with a DBH over 12 cm are measured, while on the larger, 500-m^2 plot only trees with a DBH of 35 cm or above are measured.

It is also important to consider the life span of an inventory plot when deciding on a plot size. Permanent fixed-area sample plots intended for multiple surveys are difficult to optimize. The number of trees and their size will naturally change over time. To ensure that the plot size is sufficient throughout the life of a plot, a permanent fixed-area plot tends to be relatively large. For continuous forest inventory and monitoring, fixed-area plots are to be particularly recommended, as they allow easy determination of growth components such as ingrowth, mortality, and cuts (Scott 1998). Also, fixed-area plots are usually

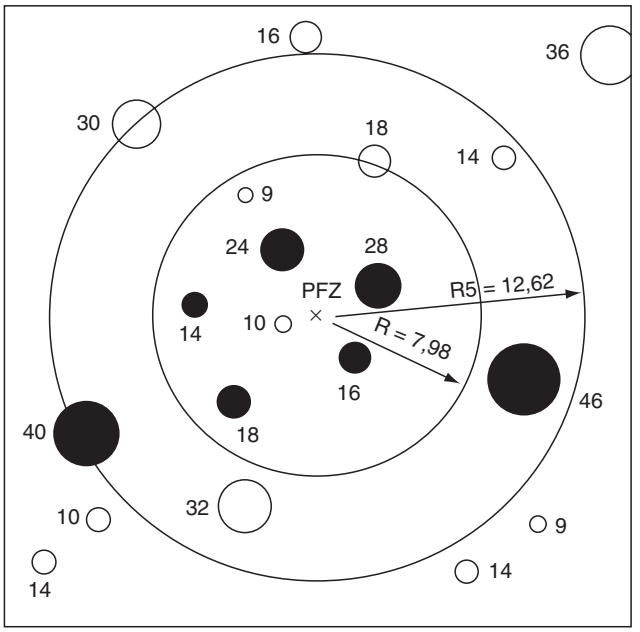


Fig. 3.12. Field plot of the Swiss National Forest Inventory (*dots* tallied trees, *circles* trees not tallied)

simple to survey, maintain, and analyze. For these reasons they are preferred over variable-radius plots.

3.5.2 Scaling of Individual Tree Data into Sample Plot Values

The statistical approach to sampling designs generally assumes that sample plots represent the smallest (natural) sample unit; however, actual sampling may not be done with this unit. Indeed plots of different size may be used or trees may be selected based on a criterion of inclusion. Thus, individual tree values sampled during a survey have to be scaled to this natural unit. It is common to take an area of 1 ha as the natural unit. The scaling is accomplished by area weighting of the attribute, say Y, of the jth tree on the ith sample location. Let a_{ij} denote the area of the sample plot used to sample the ith tree at the jth sample location. The meaning of sample plot area for trees selected by a criterion of inclusion will become clear as we later examine various sampling methods.

The area weight given to the attribute value Y_{ij} is $w_{ij} = a_{ij}^{-1}$, which becomes the attribute value per hectare. This area weighting is flexible as it extends

naturally to sampling with unequal probability of inclusion for individual trees. A simple example ought to clarify the concept. Let us assume that we are sampling with a set of fixed-area concentric sample plots. Trees below a certain size threshold are measured on the smaller plot(s) and trees above a certain threshold are measured on the bigger plot(s). Thus, the selection probability of trees is not constant but depends on tree size. The effect of different plot sizes on selection probabilities has to be corrected through scaling to a common (natural) unit via area weights. For each concentric plot a separate scaling factor applies. If, for example, two concentric plots of sizes 0.02 and 0.05 ha are used, the scaling factors are calculated as

$$w = \frac{1}{0.05} = 20 \text{ for trees on the } 0.05 - \text{ha sample plot}$$

and

$$w = \frac{1}{0.02} = 50 \text{ for trees on the } 0.02 - \text{ha sample plot.}$$

In the previous example it was the size of the trees that determined whether they were measured on one plot or the other. Often DBH is used as the size criterion owing to ease of measurement. In that case the area weights (scale factors) become functions of DBH. If in the previous example trees with a DBH between 12 and 35 cm were tallied in the smaller, 0.02-ha plot and trees with a larger DBH were tallied in the 0.05-ha plot we can express the weights as

$$w_{ij} = \begin{cases} 50 \text{ if } 12\text{cm} \leq \text{DBH}_{ij} \leq 35\text{cm} \\ 20 \text{ if } \text{DBH}_{ij} > 35\text{cm} \end{cases}.$$

Note, the scaling of inventory estimates to a unit area (here 1 ha) allows us to assess the effect of plot size, plot shape, and selection criterion on statistical estimates of interest. Recall, we do not consider a scaled attribute value as a ratio of two random variables since we assume throughout that the scale factor is known without error. Measurement errors in a_{ij} are not considered.

After scaling individual attribute values Y_{ij} to a unit area (1 ha), we usually sum them to a single value Y_{i+} for the ith sample location:

$$Y_{i+} = \sum_j w_{ij} Y_{ij}.$$

3.5.3 Point Sampling

Compared with fixed-area plots, point sampling is a relatively new procedure. It was first presented by Bitterlich (1947, 1997) in 1947 and was further refined and theoretically substantiated by Keen (1950) and then Grosenbaugh (1952). Alternative names for this method are angle count, variable plot cruising, and plotless cruising, names that reflect one of the most important features of the method: the area surveyed varies.

The first step in point sampling is the same as that in surveying with fixed-area plots: the selection of sample locations (points). Then, attributes of interest are measured on all trees meeting a certain criterion of selection. Typically the criterion is DBH and the decision on whether to include or exclude the tree from selection is based on a measurement with an angle gauge instrument. The simplest form of an angle gauge has a cross-arm attached horizontally to a vertical stick held at a known distance from the eye. With this instrument, an angle in the horizontal plane and 1.3 m aboveground is aligned with the trunk of each tree visible from the sample location. All trees with a DBH forming an angle greater than the angle subtended by the crossbar are selected. Trees with a smaller DBH are ignored (Fig. 3.13). Assessments and measurements are then carried out on the selected trees. Refinements of this method include electronic verification of inclusions based on optical/electronic measurements of angles and distances at a fixed reference height (Bitterlich 1997).

The basal area per hectare at the sample location is determined through multiplication of the number of "in" trees by a constant factor derived from the given angle subtended by the horizontal crossbar; no extra measurements are needed. Thus, each tree assessed, independent of its diameter, represents the same basal area per hectare; a proof is given next.

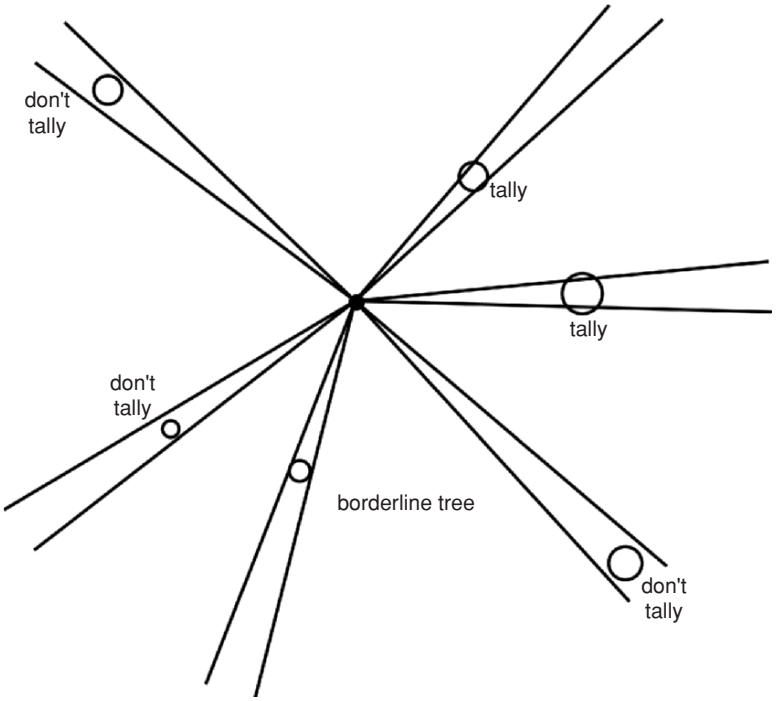


Fig.3.13. Point sampling

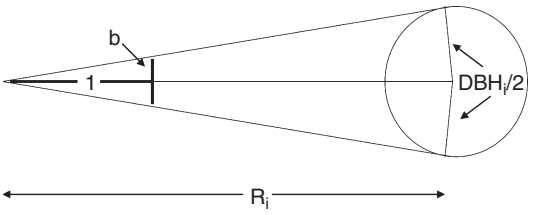


Fig. 3.14. Geometrical principle of point sampling

The geometrical principle of point sampling is illustrated in Fig. 3.14 with a cross-arm 2-cm wide (b=2 cm) attached to a stick 1-m long (l=1 m). For a tree at distance R_i with a DBH equal to DBH_i and subtending an angle equal to the angle subtended by the angle gauge cross-arm we get

$$\frac{b}{l} = \frac{2}{100} = \frac{\mathrm{DHB}_i}{R_i} \Leftrightarrow R_i = 50 \times \mathrm{DBH}_i .$$

Any sample location within a distance of $50 \times \mathrm{DBH}_i$ from this tree would result in the tree being included in the sample. In other words, the sample area for this tree is $\pi \times (50 \times \mathrm{DBH}_i)^2$. The attribute value of the tree is therefore given an area weight equal to the inverse of this area. If the attribute value is basal area $\left(\pi/4 \times \mathrm{DBH}_i^2\right)$ the area-weighted attribute value is simply

$$\frac{\pi/4 \times \mathrm{DBH}_i^2}{\pi \times \left(50 \times \mathrm{DBH}_i\right)^2} = \frac{1}{10000} \frac{\mathrm{m}^2}{\mathrm{m}^2} = 1 \frac{\mathrm{m}^2}{\mathrm{ha}}.$$

This means that every selected tree represents a basal area of 1 m^2/ha. The basal area per hectare is estimated by simply counting the number of "in" trees.

The simple derivation just shown is only valid for b/l=2/100. If a gauge with a different subtended angle (α) is used a more general equation must be employed (see Fig. 3.14 for details and definitions):

$$R_i = \frac{d_i}{2 \times \sin\left(\alpha/2\right)}.$$

For any given gauge angle α and count N_{count} of "in" trees the basal area G per hectare is estimated as

$$\hat{G} = N_{\mathrm{count}} \times 10^4 \times \sin^2\left(\alpha/2\right) = N_{\mathrm{count}} \times \mathrm{BAF},$$

where BAF is the basal area factor. The basal area factor is indicated on commercially available angle gauges. After a 360° sweep and deciding on "in"/"out" for every visible tree, one obtains the basal area per hectare by multiplying the number of "in" trees (N_{count}) by the basal area factor. The chosen angle and thus the factor determine the number of trees selected. The wider the angle, the higher the factor and the lower the number of trees selected. In tropical forests, factors between 8 and 10 are popular; they ensure reasonable counts (between 10 and 40).

As already illustrated, point sampling with an angle gauge is essentially sampling with PPS (basal area) (Fig. 3.15). In fixed-area sampling all trees have the same probability of selection, a probability that only depends on plot size. For any attribute related to size, a selection with probability of selection proportional to size will result in a more efficient sampling (Brewer and Hanif 1983; Särndal 1996). The estimated sampling error for a given number of selected trees will be lower than for sampling with equal selection probability.

It can happen that the angle subtended by a tree's DBH appears to be exactly equal to the gauge angle. Such trees are termed "borderline trees"; their diameter and distance from the point center must be measured accurately, and the decision as to whether to include them or not is based on the equation. Alternatively one could toss a coin and decide on the basis of the outcome of the coin toss.

Trees not visible from the sample location are obviously a potentially serious source of bias in point sampling with an angle gauge. Great care must therefore be taken to ensure that no tree has been missed.

Commercially available instruments for point sampling are based on one of two different principles. One uses the previously outlined principle of two divergent lines starting at the viewpoint and extending to a fixed reference distance and beyond until intercepted by an obstacle (Fig. 3.16a). The practical problem

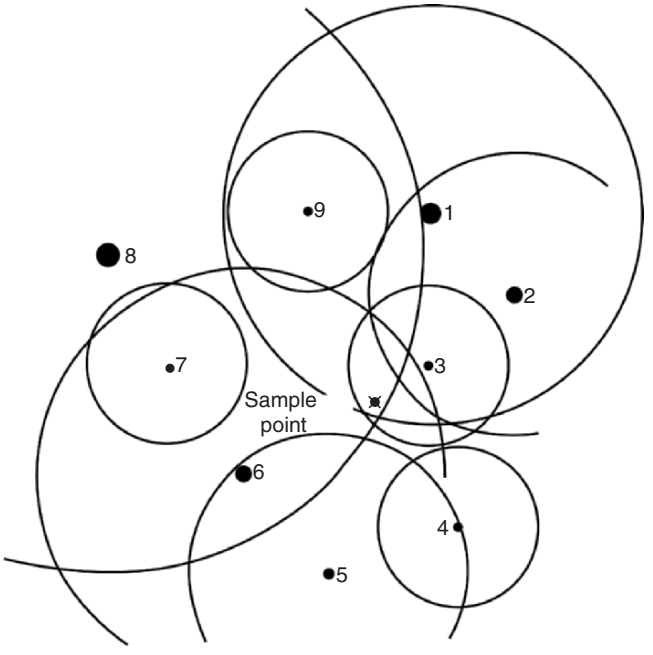


Fig. 3.15. Selection of trees in point sampling (Circles indicate the area within which a sample points needs to be located in order to select the corresponding tree)

that arises with this type of instrument is that a close object (reference distance) and a distant object (the tree) have to be focused and two lines (right and left side of the tree) observed simultaneously, often by a human observer. This renders decisions about whether or not to include borderline trees difficult. Angle gauges sold under the name of Relascope include a feature for automatic correction for inclinations from the horizontal. A wide-scale Relascope was developed for application in tropical forests (Loetsch et al. 1973).

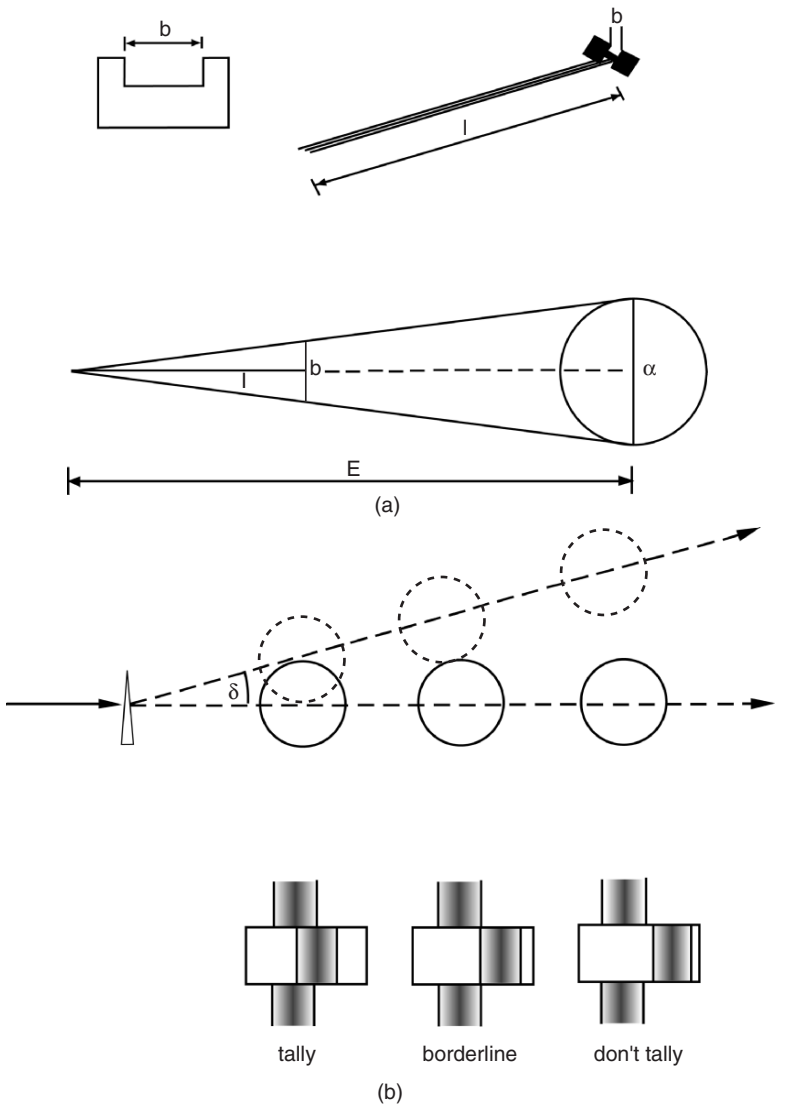


Fig. 3.16. **a** Stick-type angle gauges. **b** Wedge prisms for point sampling

The second principle is based on the diffraction of light rays, in our case diffraction of light rays from the tree as they go through a wedge prism in front of the observer (Fig. 3.16b). The observer will see two superimposed images of the tree stem: an actual nondiffracted image superimposed on a diffracted image. The tree is selected when its actual image overlaps with the diffracted image. Trees with a diffracted image displaced laterally relative to the actual image are not selected. It is much easier to decide about borderline trees with this type of instrument than with a Relascope or a stick-type angle gauge. Ease of use made them popular, especially in Canada and the USA.

With angle point sampling, any measured attribute (say population mean) of the trees counted as "in" should be expanded to a common reference area of 1 ha in order to remove the effect of differential inclusion probabilities. The expansion takes the form

$$\hat{\bar{Y}}/\text{ha} = \frac{\text{BAF}}{n} \times \sum_{i=1}^{n} \sum_{j=1}^{N_{\text{count}}(i)} \frac{y_{ij}}{\text{BA}_{ij}},$$

where BA_{ij} is the basal area of the jth "in" tree at the ith sample location. Consequently the basal area of all "in" trees must be determined or, conversely, estimated from a measurement of DBH and the assumption of a circular outline of the stem cross section. Per-hectare estimates of stems and basal area deserve special attention. For stem count, the attribute value of each "in" tree is 1, so the estimator for the stem count per hectare becomes

$$\hat{N}/\text{ha} = \frac{\text{BAF}}{n} \times \sum_{i=1}^{n} \sum_{j=1}^{N_{\text{count}}(i)} \frac{1}{\text{BA}_{ij}}$$

and the estimator for the basal area per hectare is

$$\hat{\text{BA}}/\text{ha} = \frac{\text{BAF}}{n} \times \sum_{i=1}^{n} \sum_{j=1}^{N_{\text{count}}(i)} \frac{\text{BA}_{ij}}{\text{BA}_{ij}} = \frac{\text{BAF}}{n} \times \sum_{i=1}^{n} N_{\text{count}}(i).$$

There are many variations in horizontal point sampling methodology, some have already been described. Vertical point sampling and vertical line sampling (Strand 1958), the critical height method for volume assessment and angle counting by Wenk (Loetsch et al. 1973; Hush et al. 1982) have also become popular. Uebelhör (1988) describes the use of point sampling with the wide-scale relascope in the Philippines NFI and recommends point sampling for measurements in tropical rain forests to the reduce the cost of field surveys. Other applications of point sampling in tropical forests have been presented by Boon (1970), Puffenberger (1976), da Silva (1982), and Banyard (1987). Sampling for coarse woody debris has spurred new refinements of the Relascope idea for special-purpose sampling (Ståhl 1997, 1998; Ringvall and Ståhl 1999).

3.5.4 Point Sampling Versus Fixed-Area Plots

Forest resource sampling is a challenge owing to the frequently encountered problem of uniquely defining population elements/units and consequently the problem of defining a sampling frame. The point paradigm, by which a population is defined as all possible spatial locations of a sample unit, is adopted out of necessity. When there is no natural sampling unit the survey designer has to decide on how observations are gathered at each sample location. The decision as to whether to employ point sampling or sampling with fixed-area plots depends on the individual aims and needs of the inventory. In a study on an area of 60 ha in Surinam, Schreuder et al. (1987) compared the efficiency of fixed-area plots, point sampling, and horizontal line sampling. Fixed-area plots gave the best results in terms of root-mean-square error for tree number, horizontal line sampling for basal area, the sum of tree diameters, and the average tree diameter. Point sampling was superior for estimating the number of trees in the upper-diameter classes, while fixed-area plots fared better for the smaller-diameter classes. These findings apply in general.

With point sampling in stands with a high stem count, with clusters of big and small trees, or with dense undergrowth, there is a nontrivial risk that trees may be hidden, and consequently that a negative bias may be incurred. The time to implement repeated checks for hidden trees and their inclusion in the local sample quickly erodes any practical advantage of point sampling. In such cases, it is preferable to use fixed-area plots.

In the consideration of fixed-area plots versus point sampling the survey analyst must also take into consideration the life span of a sample location. Will the sample locations be used in future inventories or will they be part of an ongoing monitoring? If plots are to be used again and again over time for the purpose of estimating change and trends in population parameters, the fixed-area plot has some distinct advantages. In point sampling, the inclusion probability of a tree depends on the attribute value of the tree (commonly basal area) at the time of sampling. Thus, the inclusion probability does not remain constant over time for attributes/variables that change over time. In the context of estimating the population parameter "tree growth" the change in inclusion probabilities generates some unique estimation problems. At the time of remeasurement you can have trees included in the point sample that were not included at the previous time of sampling. Estimation of growth of individual trees becomes cumbersome when their inclusion probability has changed between the times of measurement (Flewelling and Thomas 1984; Scott 1998).

There are two distinct events that would allow a tree to enter the later measurement but not the earlier. First, the DBH of the tree exceeded the minimum threshold diameter on the first occassion but it was located beyond the critical

inclusion radius. The growth estimate for this tree would equal its size. The common terminology for this type of growth is ongrowth. The second event that allows a tree to enter the sample between sampling times happens when its size exceeds the inclusion threshold on the second but not the first measurement occassion. The growth calculated for this tree is called ingrowth. Kuusela (1979) describes various estimators of growth based on point sampling with repeated measurements at a fixed set of sample locations. The complex nature of these estimators suggests that they should only apply in exceptional cases. Procedures for estimation of increment and growth components in fixed-area plot sampling are simple in comparison.

3.5.5 Sampling at the Forest Edge

Sample locations at the forest edge present a special estimation problem in forest inventories. Since the population of interest is restricted to areas classified as forest it can happen that the effective sample area for locations along the edge is less than the nominal area associated with sample locations in the interior of the forest. It would be wrong to simply disregard such sample locations as this would mean that trees growing in areas along the forest edge would have a different probability of being selected compared with trees growing further away from this edge (Williams 1996; Gregoire and Scott 2003). Since growing conditions and tree species along the forest edge are often distinctly different from the those in the interior forest, disregarding edge plots could lead to a serious bias in the inventory estimates.

The surveyor has several options for correctly dealing with the boundary effects. The problem and solutions are best understood if we adopt a tree-centered view of sample areas. For the fixed-area plot the sample area of a tree is simply the area covered by the sample plot when the center of the sample plot and the tree location coincide. For point sampling the sample area is the area covered by the circle centered at the tree location and with a radius equal to the critical distance of selection. Any sample location falling inside the sample area triggers the selection of the tree located in the center of the sample area. Trees located along the edge of the forest will have part of their sample area outside the population of interest. They are therefore less likely to be selected than a tree further away from the boundary. The solutions presented next are for sampling with fixed-area plots but they apply equally to point sampling by simply replacing the term sample plot by sample area. One recommended option involves finding the exact intersection of the forest edge with the inventory plot and then computing the area of the plot that is inside the population. The attribute values observed on this partial plot are scaled according to the area of the plot inside the population. The weighting scheme can also be applied to

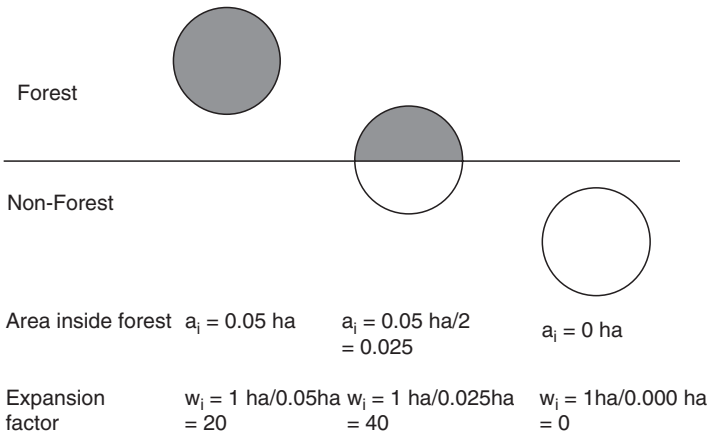


Fig.3.17. Sampling at the forest edge

trees in individual plots. For trees with a distance to the forest edge less than the radius of the appropriate sample area (e.g., 12.62 m for a circular plot of 0.05 ha) the part of the sample area inside the forest is determined for each tree and converted to an area weight w_{ij} (Fig. 3.17).

Another rather ingenious solution is the "fold-back" or the mirror reflection method (Schmid 1969). In a mirror reflection of a straddler plot, the part of the plot that falls outside the forest is projected orthogonally back into the forest with the forest edge serving as the axis of projection. The surveyor records all attributes for the part of the plot that is fully inside the forest and then all attributes on the mirror reflection of the part that was outside. In other words, a part of the plot is measured twice and occasionally three or four times if the boundary is on a corner and reflected portions of the plot overlap. Correctly applied, this method produces unbiased results, but it assumes that the forest boundary can be located accurately. In practice it is often quite difficult to decide on the exact location of a boundary. Any nonlinear edge will also generate practical problems with the mirroring. When the forest edge cannot be defined in precise and generally valid terms the method becomes problematic. Two easy, but resulting in biased estimates, solutions are to (1) relocate the straddling plot further away from the boundary to avoid overlap and (2) expand the part of the plot inside the forest to compensate for the area outside the forest. Gregoire and Scott (1990) compared four unbiased and three biased methods for dealing with sample plots at the forest edge in a mixed hardwood and mixed softwood stand in Maine, USA. They concluded that no single method was uniformly superior; the performance depends on the nature and magnitude of the "edge effect." Some biased methods of plot relocation performed, at times, better than the unbiased methods.

In practice usually only plots with a plot center inside the forest are accepted and tallied. The existence of sample locations outside the forest boundary, however, raises questions about the integrity of the sample frame and the multipliers to use to when you scale per hectare estimates to population totals. If a sample location is judged outside the forest but the outside location is actually a part of the forest estate (in an administrative or legal context) it can be argued that the "outside" location should have been included in the sample. In areas with illegal forest clearing, for example, the discarding of "outside" sample locations could lead to a serious overestimation of per hectare valued attributes.

3.6 Sampling on Successive Occasions

Sampling on successive occasions is done for the following main objectives:

- To determine the status of the forest resource at the time of the first inventory
- To determine the status of the forest resource at the time of the second inventory
- To determine changes in the forest resource between two successive inventories

The idea of quantifying change in a forest resource as the difference between two successive inventories was first applied to individual forest stands. Repeated measurements of a selected number of representative stands offered a way to verify the sustainability in terms of the yield of stands that were under a fixed forest management regime. This fundamental idea to quantify forest yield was born in the last century in Europe. In Germany, the first permanent plots were established in 1860 (Graves 1906). Foresters in France (Gurnaud 1878) developed a set of rules for how to estimate increment from successive measurement. In French-speaking countries the rules were given the name *la méthode du contrôle*. Biolley (1921) was the first to apply the rules. The forest of Couvet in the Swiss Jura, where the rules originated, was measured ten times between 1890 and 1946 in intervals every 6–7 years. The rule set has since been widely adopted. It is known as "the control method" in the English literature.

In the USA, the idea of obtaining quantitative estimates about the change in standing wood volumes through repeated measurements of the same set of plots gained support and acceptance during the years between 1929 and 1950 (Stott and Semmes 1962). The economic recession of the 1930s accentuated the need for reliable estimates of wood volume. A general increase in interest in primary production factors was instrumental in the pioneering application of sampling methods for estimating change. A direct adaptation of the European

yield control methods, which were based on recording all trees within stands up to several hectares in size, would clearly not be feasible, the intensive control method of Gurnaud (1878) and Biolley (1921) even less so. As well, there was no representative network of "benchmark" stands to which a set of standard management regimes applies. The great expanse of the North American forests, mostly without any established stand structure, dictates that only a small fraction of the forest of interest could be surveyed. The favored approach was for the application of objective and scientifically sound sampling methods, a rare approach at that time.

3.6.1 Continuous Forest Inventory

In the 1930s, a sampling method, known as continuous forest inventory (CFI), was developed in the USA. CFI is based on repeated measurements of a set of sample plots (Stott and Ryan 1939). Stott and Semmens (Stott and Semmens 1962) give a historic overview of the CFI application. In the Midwest, between 1937 and 1938, a few hundred permanent sample plots in forests operated by the wood processing industry were established. In the Great Lakes and Central Plains states starting in 1939, approximately 3,700 permanent circular sample plots were set up in private, industrial, and public forest enterprises. In 1948, the inventory of forests in Ohio and Wisconsin took place with about 1,000 permanent sample plots. In 1952, the American Pulpwood Association (APA) became aware of the CFI and introduced it to its members. During the following years, a cooperation between the APA and the USDA Forest Service led to an extensive application of the CFI extending east of the Mississippi River. In 1962, approximately 50 enterprises associated with the wood processing industry managed 25 million acres according to the CFI method. Most CFI plots were established in what was termed "typical" timber-producing stands; as such they are not representative of the entire forest resource.

The pioneers of CFI in Germany were Krutzsch and Loetsch (1938). In 1936 they set up a series of permanent sample plots for yield control. In Sweden, CFI was pioneered by Patterson (1950) and early on applied to forest yield research at the Swedish forest experimental station. In Switzerland, CFI was introduced by Schmid (1967) and it was applied to forest management planning, in effect an extension of the classic control method to CFI. His intensive effort towards an applied survey method for permanent sample plots (Schmid-Haas et al. 1993) resulted in wide acceptance of the method in Swiss forestry.

With the CFI method, all sample plots, which are measured on the first occasion, are measured again in successive inventories. The estimators of population parameters under CFI are time-specific. We indicate the time dependency of

CFI estimators by a suffix t for time. The suffix takes on values 1,2,3, . . . for the estimates of population parameters at the first, second, third, and so on inventory. We are usually interested in the estimation of a change between two successive inventories. When the context is clear we simply refer to estimators and estimates of the "first" occasion and of the "second" occasion inventory, respectively. In continuation of our example for the estimation of a population mean under SRS at the first and the second occasion inventory, the CFI estimator is

$$\hat{\bar{Y}}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} Y_{t,i}, t = 1,2.$$

Changes in a population parameter between two inventories can be derived as the difference between the estimates of the population parameter at two successive inventories. For the previous example we have

$$\Delta\hat{\bar{Y}}_{2,1} = \hat{\bar{Y}}_2 - \hat{\bar{Y}}_1.$$

When the same set of plots are remeasured on both occassions, the estimator of the variance of the change of the mean becomes

$$\hat{\text{var}}\left(\Delta\hat{\bar{Y}}_{2,1}\right) = \hat{\text{var}}\left(\Delta\hat{\bar{Y}}_2\right) + \hat{\text{var}}\left(\Delta\hat{\bar{Y}}_1\right) - 2 \times \hat{\rho}\left(Y_2, Y_1\right) \times \sqrt{\hat{\text{var}}\left(\hat{\bar{Y}}_2\right)\hat{\text{var}}\left(\hat{\bar{Y}}_1\right)},$$

where $\hat{\rho}\left(Y_2, Y_1\right)$ is an estimate of the correlation coefficient between the observations on the second and the first occasion; $\hat{\rho}\left(Y_2, Y_1\right)$ is restricted to values between −1 and +1.

The higher the correlation is between paired observations from the first and second inventory, the smaller is the variance of their difference. For a large number of size-related attributes the temporal correlation between plot variables measured on two occasions will be positive. Autocorrelation is the single most significant contributor to this correlation (the attribute on the second occasion is equal to the attribute on the first occasion plus change). When the correlation is positive the variance of the change will be less than the sum of the variances on the first and second occasions. However, as time separates the two inventories the correlation tends to dissipate. The rate of decrease will depend on how well change is correlated with the attribute value on the first occasion. For trees growing in the absence of disturbances, in a homogenous environment free of competition, and with a nonrestrictive supply of nutrients, the correlation between change and initial attribute value may be quite strong over long periods of time. In heterogeneous environments with frequent disturbances and physiological stress the correlation may be weak, zero, or even negative. When the correlation is zero or perhaps even negative the variance of the difference will be equal to or larger than the sum of the respective variances. For example, if large trees are more prone to hurricane damage than small trees, and one or more hurricanes have gone over the forest since the previous inventory, the correlation between, say, plot volume at the two occasions could be negative. At the other extreme, when the correlation is 1, as is the case when the attribute value

on the second occasion is equal to the attribute value on the first occasion times a constant plus a constant, the variance of the change is 0. A correlation coefficient of 1 is extremely rare. If the plots measured on the first and the second occasion are not the same and each set is selected independently of the other then the correlation of attribute values is by definition zero and the variance of the change estimate is simply the sum of the respective variances.

For a positive correlation coefficient a CFI estimate of change will have a smaller variance than an estimate change derived from two sets of independent observations. The advantage of using the CFI method rests with the reduction of the variance of estimated change.

The CFI method, despite its obvious advantage, encounters practical and inferential problems. Over time the locations of sample plots may become known beyond the surveyors and, as a result, they may evolve differently from the surrounding forest. This nontrivial risk is especially acute for visibly marked sample plots. The potential of an inferential problem is latent because, as paraphrased by Schmid-Haas (1983), "there is no guarantee that sample plots, visible or not, will remain representative of the target population." Schmid-Haas also believes that even the most experienced forester cannot be sure that he or she would not be influenced by the knowledge that certain parts of the forest are subject to the intensive scrutiny of repeated measurements. Consciously or unconsciously, it is possible that the sample locations are being treated differently in some way, shape, or form. A sample plot inventory, which cannot reliably eliminate this risk, may become biased and will quickly lose credibility and invested goodwill.

If only the net change has to be estimated, for example, volume growth, permanent sample plots would be more cost efficient than two independent surveys, which means that for the same cost they lead to a smaller sample error. This seems obvious, since the difference between two independent observations is not only caused by change alone, but also by the variation within the two populations. If only the current state is to be considered, temporary sample plots are often more cost effective than permanent plots, since the expenditures for marking the sample plot centers and the registration of sample tree locations do not exist.

The application of the CFI method can lead to inferential problems. All CFI inventory systems rely on the assumption that the permanent plots are representative. But are they? With time the plots may "drift" at a rate different from that of the population they are supposed to represent. This risk is especially acute in managed forests or in places with frequent land-use changes. As well, changes in the inventory objectives are difficult to accommodate in CFI with its system of plots established in the past and tailored to past objectives.

Practical survey objectives are often a blend of target precision on estimates of state and change. In this case a design with a mixture of permanent and

temporary plots appears attractive. The idea of a survey design with both plot types arose from these considerations.

3.6.2 Sampling with Partial Replacement of Sample Plots

In SPR a fraction of the sample plots measured during a survey are replaced by new sample plots at the subsequent survey. This pattern of partial replacement is repeated over time. SPR was introduced into forest inventory around 1960. Kish (1964), Cochran (1977), and Sukhatme et al. (1984) also discussed the theory of SPR of sample plots. Bickford (1959) was the first to introduce the theory of SPR to forest inventory. The first to apply SPR was the USDA Forest Service in the Allegheny National Forest where it was combined with aerial photographs and modified accordingly (Bickford 1959).

We shall describe SPR estimators based on only two SPR occasions for the estimation of a population mean. SPR estimators for subsequent surveys are more complex. After the second SPR there are three types of plots available for the estimators:

1. n_{12} sample plots measured on both occasions (matched permanent plots)
2. n_1 sample plots measured only on the first occasion (unmatched first occasion plots)
3. n_2 sample plots measured only on the second occasion (unmatched second occassion plots)

The most precise unbiased linear estimator of the state on the first occasion $\hat{\bar{Y}}_1$, on the second occasion $\hat{\bar{Y}}_{2|1}$, and of change between the two occasions $\Delta\hat{\bar{Y}}_{21}$ is

$$\hat{\bar{Y}}_1 = \frac{\sum_{i=1}^{n_1} Y_{i1} + \sum_{j=1}^{n_{12}} Y_{j1|2}}{n_1 + n_{12}} = \frac{n_1}{n_1 + n_{12}}\hat{\bar{Y}}_1 + \frac{n_1}{n_1 + n_{12}}\hat{\bar{Y}}_{1|2},$$

where $Y_{j1|2}$ is used to denote the attribute value on the first occasion on the n_{12} matched plots. A corresponding estimator for the second occasion $\hat{\bar{Y}}_2$ is obtained by a simple switch of occasion subscripts. The best estimator of the status on the second occasion exploits the relationship between the attribute values on the first and the second occasion:

$$\hat{\bar{Y}}_{2|1} = \hat{c} \times \left[\hat{\bar{Y}}_{2|1} + \hat{\beta}_{2.1}\left(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_{1|2}\right)\right] + (1 - \hat{c})\hat{\bar{Y}}_2,$$

where $\hat{\beta}_{2.1}$ is the ordinary least-squares regression coefficient obtained by regression of $\hat{\bar{Y}}_{2|1}$ on $\hat{\bar{Y}}_{j1|2}$ and $\hat{c}$ is an estimate of the optimal weight to be assigned to the first term, which is essentially the estimator used for double sampling with regression estimator. The optimal weight is

$$\hat{c} = \frac{n_{12}(n_1 + n_{12})}{(n_1 + n_{12})(n_2 + n_{12}) - \hat{\rho}_{2.1}^2 \times n_1 \times n_2},$$

where $\hat{\rho}_{2.1}$ is an estimate of the correlation coefficient between the second and the first occasion attribute values. The best unbiased linear estimator of change is

$$\Delta\hat{\bar{Y}}_{21} = \hat{\bar{Y}}_{2|1} - \hat{\bar{Y}}_{1|2},$$

where $\hat{\bar{Y}}_{1|2}$ is the double sampling with regression estimator of status on the first occasion computed as

$$\hat{\bar{Y}}_{1|2} = \hat{e}\left[\hat{\bar{Y}}_{1|2} + \hat{\beta}_{1.2} \times \frac{n_2}{n_{12} + n_2}\left(\hat{\bar{Y}}_2 - \hat{\bar{Y}}_{2|1}\right)\right] + (1 - \hat{e}) \times \hat{\bar{Y}}_1,$$

with

$$\hat{e} = \frac{n_2 + n_{12}}{n_1 + n_{12}} \times \hat{c}.$$

The variance estimator for $\hat{\bar{Y}}_1$ is obtained as for SRS, while that of $\hat{\bar{Y}}_{2|1}$ is

$$\hat{\text{var}}\left(\hat{\bar{Y}}_{2|1}\right) = \frac{\left(n_1 + n_{12} - \hat{\rho}_{2.1}^2 \times n_1\right) \times \left[\sum_{i=1}^{n_2}\left(Y_{i2} - \hat{\bar{Y}}_2\right)^2 + \sum_{j=1}^{n_{12}}\left(Y_{j2|1} - \hat{\bar{Y}}_2\right)^2\right]}{\left[(n_1 + n_{12})(n_2 + n_{12}) - \hat{\rho}_{2.1}^2 \times n_1 \times n_2\right](n_2 + n_{12} - 1)}.$$

Finally, an unbiased (but not minimum-variance) SPR estimator of change is

$$\hat{\text{var}}\left(\Delta\hat{\bar{Y}}_{21}\right) = \left(1 - \hat{k}_2\right)\frac{\hat{\text{var}}\left(Y_2\right)}{n_2} + \left(1 - l^2\right)\frac{\hat{\text{var}}\left(Y_1\right)}{n_1} + \frac{\hat{k}^2\,\hat{\text{var}}\left(Y_{2|1}\right) + l^2\,\hat{\text{var}}\left(Y_{1|2}\right) - 2\hat{k} \times l\text{cov}\left(Y_{1|2}, Y_{2|1}\right)}{n_{12}}.$$

where

$$\hat{k} = \frac{n_{12}(n_1 + (n_2 - n_{12})\hat{\beta}_{2.1})}{n_1 n_2 - \hat{\rho}_{2.1}^2(n_1 - n_{12}(n_2 - n_{12})}$$ and

$$\hat{l} = \frac{n_{12}(n_2 + (n_1 - n_{12})\hat{\beta}_{2.1})}{n_1 n_2 - \hat{\rho}_{2.1}^2(n_1 - n_{12}(n_2 - n_{12})}$$

Ware (1960) examined inventory data from repeated measurements in the northeastern region of the USA and found that in six out of eight cases, the variance was not the same on both inventory occasions. It is therefore important not to simplify the change estimator by assuming equal variances if they are not. Violations of this assumption results in a biased estimator.

Ware and Cunia (1962) championed for a wider use of SPR in forest inventories. SPR at that time was mainly of theoretical interest and practical applications were few. Optimality of SPR for change estimation requires either the equality of population variance or the same sample size in successive inventories, or both. An optimal rate of replacement of sample units is only solvable for

a single attribute. The problem became intractable for multivariate attributes. Different survey costs of new and repeatedly measured plots further increase the complexity of an already complex optimization problem.

Scott (1981, 1984) provided a similar set of estimators derived from the work by Meier (1953). Scott and Köhl (1994) applied the SPR concept to two-phase sampling for stratification on two and three occasions.

After more than two inventory occasions, the best SPR estimator becomes very complex and unwieldy (Scott 1986; Scott and Köhl 1994) thanks to a myriad of plot types and pairwise associations between plot measurement values exploited in the estimators. With only two inventory occasions, we have three different types of plots. With three inventory occasions, there would be a six types of plots with sample sizes n_1, n_2, n_3, n_{12}, n_{123}, and n_{23}. With four inventory occasions, we would have ten plot types. Therefore, the complexity increases rapidly with the number of observations over time. As design imbalance is also bound to creep in over time; even the most ambitious SPR design can barely stand the test of time.

The problems encountered in practical implementation with SPR are clear detractors. In some survey regions of the USA, SPR has recently been replaced by more flexible and less complex designs (Hahn and Scott 2003, personal communication) such as, for example, a semisystematic sampling design where plot location is random within a regular tesselation of the population into equal-sized hexagons.

3.6.3 Estimates for Subpopulations

Inventory results are not only needed for the entire population, but frequently also for thematic subunits, such as, for example, the forest area structured by ownership categories, by site quality, or by forest cover type. A tabular representation of subpopulation estimates arranged in one-way, two-way, or multi-way tables accommodates this need. The margins of these tables provide row, or column, totals of one or more thematic subunits. When cell and marginal estimates are obtained independently of each other, the additivity of estimates is no longer assured. It will depend on the sampling design and the estimators used to obtain estimates for individual cells. Only SRS and CFI estimators result in additive tables (Table 3.3). Two-phase sampling and SPR designs are

Table 3.3. Example of an additive table. Forest area by type of ownership and site quality in 1,000 ha

	Poor/moderate	Good/very good	Total
Public forest	404.1	408.0	812.1
Private forest	114.5	259.7	374.2
Total	518.6	667.7	1,186.3

Source EAFV (1988, p. 81)

Table 3.4. Example of a nonadditive table. Forest area by type of ownership and site quality in 1,000 ha

	Poor/moderate	Good/very good	Total
Public forest	409.1	407.0	824.9
Private forest	119.4	256.8	370.3
Total	503.1	671.9	1,186.3

notorious for creating nonadditive summary tables. For all other designs and estimators the cell values may not add to the row, or column, sums. An example is given in Table 3.4. Nonadditive tables are not a problem per se for a statistician. Nevertheless, we can hardly expect users of inventory results to accept nonadditive tabulated results. Consequently, a need to remove the nonadditivity comes around frequently to the inventory analyst. The most popular method is based on variants of iterative proportional fitting in which the row, or the column, discrepancies are distributed across cells in proportion to their row, or column, sums (Bishop et al. 1975; Li and Schreuder 1985; Zhang and Chambers 2004). Another popular approach computes EB posterior predictive estimates based on a model for the entire table (Green et al. 1992, Laird 1978). These model-based estimates are, by definition, additive.

3.7 Sampling for Rare and Elusive Populations

Sampling for estimation of population totals, density, or the total or mean attribute value when the population is rare (elusive) will require a large sample size to get the sampling error down to an acceptable level. Efficient sampling becomes paramount in order to control the cost and time needed to reach a target precision. Exploiting auxiliary information associated with the population of interest becomes especially attractive in this situation. Knowledge about the spatial distribution of the population can also improve the efficiency of sampling by choosing a design that is specifically tailored to do well under the assumed distribution (Kalton and Anderson 1986; Sudman et al. 1988; Christman 2000; Venette et al. 2002).

One of the real enigmas in sampling a rare/elusive population is the risk of an empty sample. To state that the estimated population total is zero with a sampling variance of zero (all sample values are zero) is a very strong statement that is bound to attract a lot of attention. Consider the interest in rare population and our concern about the disappearance of species and it is easy to fathom the questions that may flow from an estimate of zero. The surveyor can guard against a zero sample if one has a prior estimate p_0 of the probability that a sample unit will contain a population element. The chance of not having a

zero sample if n units are sampled independently is then $(1 - p_0)^n$. Let us say that p_0 is estimated at 0.06 and that the surveyor wishes to keep the probability of an empty sample to 0.01 or less. A sample size equal to or greater than 765 would be needed.

Often sample sizes of this magnitude cannot be afforded and the surveyor may end up with an empty sample. The notion of a zero sampling variance is formally correct but clearly untenable in practice unless one is positive that the population has disappeared, but then why have a survey? Instead of a zero variance we suggest deriving a sampling error based on the rule-of-three (Jovanovic and Levy 1997). The rule states that $3/n$ is the upper 95% confidence bound for a binomial probability when in n independent trials no event has occurred. Since a $(1-\alpha)100\%$ upper bound (p) for a binomial probability can be found by solving $(1-p)^n \geq \alpha$ for p (n is fixed by design) with the approximate solution $p_{1-\alpha} \approx -\log\alpha \times n^{-1}$. For an upper 95% bound we get $-\log 0.05 = 2.996 \approx 3$, which is the essence of the rule-of-three. Assuming a half-normal distribution of sampling errors the approximate standard error of the zero estimate according to the rule-of-three would be $0.9227 \times n^{-1.5}$. To reach this result we first found the scale parameter of a half-normal distribution with a 95% quantile equal to $3/n$ and then obtained the standard deviation in a half-normal distribution with this scale parameter.

We shall now turn to potentially suitable design options for the sampling of rare/elusive populations.

3.7.1 Adaptive Cluster Sampling

Many rare and elusive populations are naturally clustered in space. The density of population elements can be quite high in a few scattered locations with favorable conditions. In surveys of rare/elusive populations the costs associated with traveling from one plot to another are often several orders of magnitude higher than the cost of measuring a plot. It would therefore seem to make sense to intensify sampling in areas with positive finds of the rare elements and reduce the time spent at plots with zero finds. Thompson (1990) has devised an adaptive design that achieves this goal. In adaptive designs the sampling effort is intensified at sample locations with a positive find. Strict adherence to a set of rules governing how and when sampling is intensified allows design-unbiased estimates of population attributes and their sample variances (Thompson 1990). There are many innovative adaptive designs for stratified (Thompson 1991), two-stage (Thompson 1991; Salehi and Seber 1997; Muttlak and Khan 2002; Smith et al. 2003), systematic (Acharya et al. 2000), line (Morgan 1997), restricted (Lo et al. 1997; Brown and Manly 1998; Christman 2001; Muttlak and Khan 2002), and point sampling (Roesch 1993).

Most of these designs can be integrated into one of the more conventional survey designs presented throughout this book. We shall take a closer look at adaptive cluster sampling as it is probably the one with the greatest potential for application in forest inventories.

In adaptive cluster sampling an initial simple random sample is taken with n fixed-area plots composed of one or more basic units of a fixed size and a geometry that completely tessellates the population. The population is viewed as composed of a regular array of N basic units. When a sample plot contains one or more of the desired population elements, the area surrounding the plot is searched for additional occurrence of elements. Around each plot with a positive find one imagines a regular grid of basic units with the actual plot located at the center. All surrounding units with a positive attribute value and connected directly or indirectly to the plot with a positive attribute value are included in the sample. All empty units along the outside edge of included nonempty units are also included, but only nominally because they have to be searched in order to determine whether they should be included or not. Two plots are directly connected if they share a common side and indirectly if they are connected through an unbroken chain of connected plots. In other words, an entire cluster of plots containing one or more population elements is included in the sample whenever the initial sample intersects a cluster. With this protocol more than one spatial cluster of elements may be included in the sample at a single sample location if the clusters are connected at the scale of the basic unit. Conversely, different sample plots may intercept the same cluster of connected units. The size and geometry of the basic unit will thus influence the connectivity of clusters and ultimately the efficiency of adaptive cluster sampling. The example given next will illustrate and clarify the sampling protocol.

To appreciate the design-unbiased sample estimators for this type of adaptive sampling it is helpful to view the population as composed of networks (Ψ) of basic units. There are two types of networks. One is made up of all empty units. Networks of this type are all of size $x = 1$ and have an attribute value $y = 0$. The other type of network consists of a set of connected (directly or indirectly) basic units each with one or more population elements. There are a finite number of networks in the population, say K, with sizes x_i, $i = 1,\ldots,K$ in basic units and attribute values of y_i, $i = 1,\ldots,K$. Thus, adaptive cluster sampling can be viewed as a SRS of networks and their associated attribute values (x and y). Our initial sample serves to intercept the networks. For plots composed of several basic units the adaptive sampling protocol is applied to each unit in the plot. The effect of multiunit sampling is in the number and mixture of intercepted networks. Alternative definitions of a network and connectivity are possible (Christman 2000), but the one just given is usually preferred and is the simplest to implement in the field.

A modified design-unbiased Hansen–Hurwitz (HH) estimator of the population mean under adaptive cluster sampling is (Thompson 1990)

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_{\psi_i}}\sum_{j\in\psi_i}Y_j,$$

where Y_j is the attribute total in the jth sample unit in the network(s) intercepted by the ith sample plot and X_{ψ_i} is the size in basic units of the network(s) intercepted by the ith sample plot. Empty networks of size 1 will dominate the sum when sampling rare/elusive populations. Note also that the empty edge units included in the sample do not appear in the estimate of the mean. In modified HH estimators the edge units are basically irrelevant. They do affect effective sample size since all units on the outside edge of an intercepted or connected nonempty unit have to be checked for inclusion or not.

The fact that only the average per unit attribute value of intercepted network(s) enters in the modified HH estimator of the population mean illustrates that the within-network variance of Y_j does not contribute to the sampling variance of the estimated mean. All sample plots intercepting a network(s) will record the attribute for that network(s). In SRS, recordings are strictly on a per plot basis. When the initial random sampling is without replacement the variance estimator of the population mean is

$$\hat{v}ar\left(\hat{\bar{Y}}\right) = \frac{(N-n)}{Nn(n-1)}\times\sum_{i=1}^{n}\left(\hat{\bar{Y}}_i - \hat{\bar{Y}}\right)^2,$$

where N is population size in basic units and $\hat{\bar{Y}}_i$ is the mean attribute value per unit in the network(s) intercepted by the ith sample plot. The modified HH estimators of the population mean and sampling variance are basically identical to the corresponding estimators under SRS. Again, if one considers adaptive cluster sampling as a sampling of networks the similarities are to be expected. Horwitz–Thompson estimators are more complex since they involve computing the inclusion probabilities of edge-units. Since edge units have to be searched they are viewed as part of the sample. Edge units can be selected if they are either intercepted by a sample plot or because they are an edge to one or more networks intercepted by the initial sample. Horwitz–Thompson estimators are said to be less sensitive to the spatial distribution of population elements than the HH estimators (Salehi 1999, 2003; Christman 2000, 2002; Felix-Medina 2003). In adaptive cluster sampling the number of basic units included in the final sample is at least the same as in the initial random sample. The expected sampling variances in adaptive cluster sampling are therefore less than the expected sampling variance in SRS. This argument extends naturally to stratified adaptive sampling. If the population is truly clustered with a sizeable within-cluster variation of the attribute of interest and the extra costs associated with delineating and searching for networks are modest compared with the cost of interplot travel then the adaptive sampling approach can be very efficient (Christman 2002; Brown 2003). It should be mentioned though that the efficiency can fluctuate widely as a function of the spatial distribution of attribute values (Acharya et al. 2000; Hanselman et al. 2003; Smith et al. 2003).

An example of adaptive cluster sampling will hopefully clarify the sampling protocol and the estimators. A survey to determine the number of rare orchids within a 110.2-ha forested area is to be conducted. From ecological studies of the orchids we know that they are clustered in a few suitable locations scattered throughout the forest. Most orchid clusters have 11–27 individuals with a mean cluster size of 16 and a standard deviation of 3.8. A cluster occupies between 25 and 270 m^2 (median 200 m^2). It is estimated that the orchids occupy less than 1% of the forested area. The survey designer decided that adaptive cluster sampling with square 5-m×5-m plots would be a suitable approach. The sampling frame can be regarded as 44,531 basic units of 5 m×5 m located in a regular grid. According to the prior information about the orchids the expected network sizes would be between one and eleven 5-m×5-m units. Differential GPS will be used to stake out sample plots and the networks of connected 5-m×5-m units with an accuracy that warrants the assumption of no errors in the orchid counts. To safeguard with 99% probability against the possibility of an empty sample (see before) a sampling intensity of 2% or 891 is deemed adequate. If the orchid counts of 5-m×5-m units are distributed approximately as a Poisson distribution with a mean of $\lfloor 0.01 \times 25 \rfloor$ orchids per unit we would expect a relative standard error of 15% on the estimated population size. We introduce this detail to highlight that sampling for rare and elusive populations is a costly endeavor. What is not known to the surveyor but is listed here for the sake of completeness is that there are a total of 480 orchids in the population (equivalent to 4.4 per hectare) distributed across 31 networks occupying a total of 243 5-m×5-m units (0.56% of the total). A map of the orchid clusters is shown in Fig. 3.18.

In the initial SRS, a total of 13 orchids were found in four plots, whereas the remaining 887 plots had no orchids. The orchid counts in the six nonempty plots were 4, 1, 4, and 4. Without adaptive sampling around these nonempty plots the estimated population density would have been 5.8 per hectare with an estimated error of 2.7 per hectare. The adaptive search of networks around the four nonempty plots added a total of 27 new units to the sample, i.e., a total of 918 units were sampled. Orchid counts in the four networks were 8, 8, 8, and 7, and the sizes of these networks were 21, 13, 15, and 18 units (of 25 m^2). Inserting these figures in the previous modified HH estimators yields a density estimate of 3.9 orchids per hectare with a standard error of 2.0. For the extra effort of delineating and counting units in six networks we have obtained estimators that are clearly superior to what we would have obtained had we stuck with a SRS design. Four of the six networks intercepted by the initial sample are shown in Fig. 3.19. Edge units have not been highlighted, but there would be eight, ten, eight, and ten edge units surrounding the four intercepted networks. Note that three of the four networks joined one or more additional networks at a plot corner. By the adopted definition of connectedness they are not to be included in the sample.

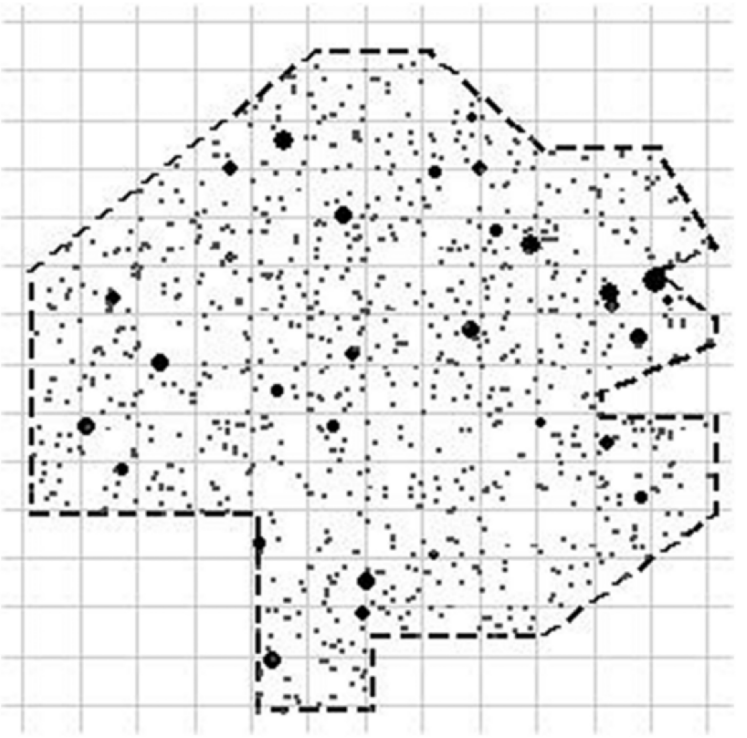


Fig. 3.18. Map of 31 orchid cluster locations (*dark circles*). Cluster sizes are proportional to the number of orchids in a cluster. Sample locations of 875 5-m×5-m sample plots are indicated by *light-gray squares*. The grid spacing is 100 m×100 m

We stated before that the efficiency of adaptive cluster sampling in terms of sampling variance relative to that of a SRS depends on the within-cluster variance of orchid density. For a fixed among-unit variance of orchid density the among-network variance in orchid density declines as the within-network variance increases and vice versa. In the previous example the within-network variance of orchid density was approximately 1.6 times as large as the among-network variance.

3.7.2 Sampling with Probability Proportional to Size

A concentration of sampling efforts in locations with a higher density of rare/elusive population units has a substantial potential for boosting the efficiency of sampling. Sampling with unequal probability is designed to give a

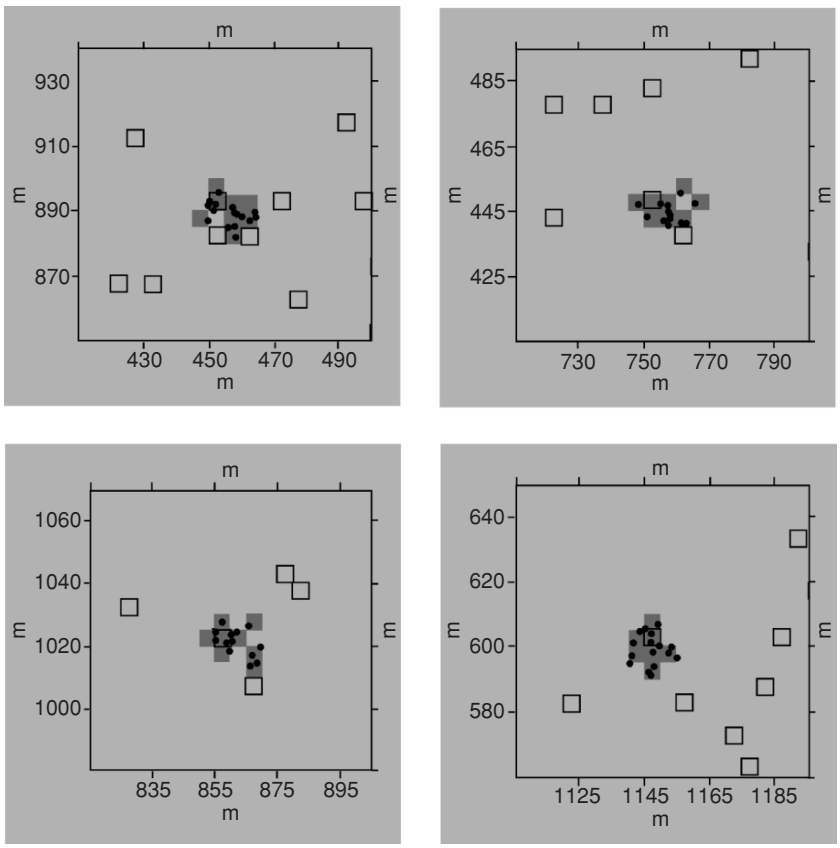


Fig. 3.19. Four intercepted networks of connected sample units with orchids (*dark-gray squares*). Initial random sample plots are indicated by *open squares*. Orchids are displayed as *black dots*

higher probability to sample locations where one expects the highest return (Brewer and Hanif 1983; Godambe and Thompson 1988; Särndal 1996). The challenge is to find an auxiliary attribute that is both known for the entire population and also is approximately proportional to the attribute of interest. Point sampling with an angle gauge is but one example of sampling attributes of trees with probability proportional to basal area. Other applications include, for example, volume sampling from known sample tree lists of basal areas (Schreuder et al. 1968, 1971, 1992; Gregoire and Valentine 1999; Magnussen 2000). Classified remotely sensed images of a population may also provide clues about the location and quantities of interest which can be used in the process of selecting samples (Ståhl et al. 2000; Williams 2001). The general

theory and estimators for sampling with unequal probability can be found in many forestry textbooks, for example, de Vries (1986), Schreuder et al. (1993), and Shiver and Borders (1996). Estimators for sampling with and without replacement are distinctly different and the latter are usually quite complicated for all but the simplest cases (Brewer 2000) owing to nonnegative joint inclusion probabilities. Although sampling with replacement is less efficient than sampling without replacement owing to the potential of repeat samples with no new information, the computational challenges involved in the estimators of variance are such that we shall forgo this efficiency and present only the estimators for sampling with replacement.

Let x_i be the auxiliary attribute and y_i the attribute of interest for the ith population unit; x_i is known for all units in the population. The sum of x_i over the entire population is T_x. We are interested in estimating the population total Y. The draw-by-draw inclusion probability of the ith population unit is $\pi_i = n \times x_i \times T_x^{-1}$ with n equal to the desired sample size. When the sample is selected at random with these inclusion probabilities the unbiased Horvitz–Thompson estimator of the population total is (Brewer and Hanif 1983)

$$\hat{Y}_{\text{PPS}} = \sum_{i \in s} y_i \times \pi_i^{-1},$$

where the summation is over the units in the sample (s). The unbiased sample-based estimator of sampling variance is

$$\hat{\text{var}}\left(\hat{Y}_{\text{PPS}}\right) = \frac{1}{n \times (n-1)} \sum_{i \in s} \left(\frac{y_i}{\pi_i} - \hat{Y}_{\text{PPS}} \right)^2.$$

We illustrate the PPS methods using the same population of orchids as exemplified under adaptive cluster sampling. We assume that the orchids are typically associated with a certain group of tree species and that this group of species can be identified with reasonable success by interpreters of remotely sensed images. A classified grayscale remote-sensing image of the forest with $5 \times 10^4 \times 5 \times 10^4$ pixels is in Fig. 3.20. The grayscale levels are assumed to reflect the likelihood of the tree species group being associated with orchids. A darker tone reflects a higher belief in the occurrence of orchids and vice versa. Pixel values were generated synthetically from the following algorithm:

$$x_i = 0.5 \times \left[\log\left(y_i + 1\right) + \gamma_i + \frac{1}{4} \sum_{j \sim i} \log\left(y_j + 1\right) + \gamma_j \right],$$

where

$$\gamma_i \sim \Gamma\left(0.4, 1\right),\ E\left(\gamma_i\right) = \text{var}\left(\gamma_i\right) = 0.4$$

and where the summation is over the four first-order neighbors to pixel j (to mimic scaling and sensor spread function). The average signal-to-noise ratio was 0.15 and the correlation between the feature of interest and the grayscale value was 0.16.

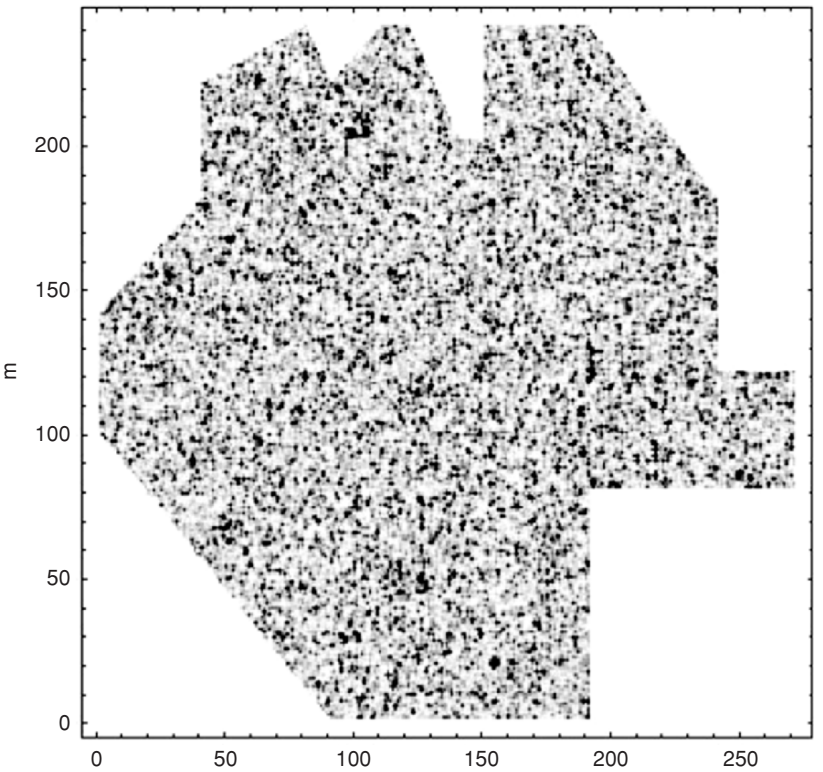


Fig. 3.20. Grayscale map of the spatial domain of the orchid population. A darker tone reflects a higher belief in the presence of orchids

Two 15-pixel×15-pixel windows taken from the grayscale map reveal the clustering of darker pixels with the presumed highest likelihood of orchid presence (Fig. 3.21).

The sample size of 891 was maintained and each unit was selected by forming a two-column list with the number (1, ...,44,531) of each population unit in the first column and the cumulative inclusion probabilities of these units in the second column. Now draw 891 random numbers uniformly distributed between 0 and 1. For each random number, find the unit number of the first cumulative inclusion probability that is larger than or equal to the random number. Select the population unit associated with that number. This selection protocol ensures that units are selected with probability proportional to π_i. For example, you have drawn the random number 0.447297. Excerpts of the numbered list of cumulative selection probabilities are in Table 3.5. The highlighted population unit with the number 19,839 should be selected.

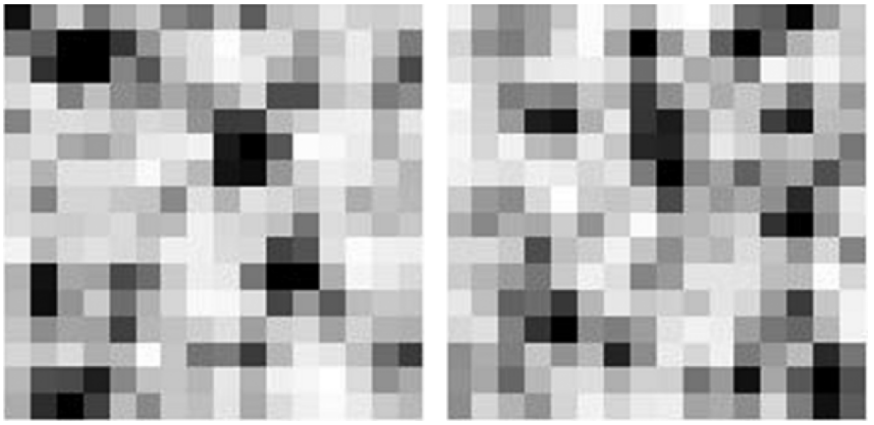


Fig. 3.21. Two randomly chosen 15-pixel×15-pixel windows of details taken from the grayscale map in Fig. 3.20

After selection we computed the average inclusion probability of the sample: it was 1.8 times higher than the average inclusion probability of a unit. The grayscale values of selected units were also slightly more strongly correlated with the actual orchid count (0.18) than seen across the entire population (0.16). A total of 30 orchids were recorded for the PPS sample (compare with 13 for SRS and 31 for the adaptive cluster sampling design). The PPS estimator of orchid density was 4.3 (actual is 4.4) with a standard error of 1.4, a clear improvement compared with the aforementioned alternatives. This PPS example was governed by realistic choices of the correlation and association between the auxiliary and the target attribute. Even with a modest relationship the possible gains in efficiency are striking. A note of caution is nevertheless warranted. If the assumptions about the auxiliary variable turn out to be

Table 3.5. Selection list of unit numbers and cumulative draw-by-draw inclusion probabilities. Random draw is 0.447297 and unit 19,839 should be selected

Unit no.	$Sum\left(\pi_{no.}, 1, no.\right)$
1	7.8×10^{-6}
⋮	⋮
19,837	0.447272
19,838	0.447282
19,839	**0.447303**
⋮	⋮
44,530	0.999949
44,531	1

wrong, or the association is very heterogeneous across the population, a PPS sample may perform poorer than a simple random sample.

3.7.3 Line Transect Sampling

As the name implies, the survey is conducted along one or more survey lines. In a survey of mobile population elements like animals, birds, and insects the population is assumed fixed in size and location of each population element during the time of the survey. How realistic this assumption is depends naturally on the time and manner in which the survey is conducted. For immobile elements this assumption is implicit and mostly quite reasonable. Line transect sampling does not depend on the existence of a sample frame, a feature it shares with point sampling designs. In a line transect an observer moves along the transect line(s) in an nonintrusive manner and records sightings of population elements and their distance to the survey line. The line can be staked out on the ground or can be a line on a photograph or some other medium and the observer can move on foot, in a vehicle, or in some elevated platform of observation, for example, an airplane. It is a strict requirement that no element is recorded more than once and that observations are mutually independent. That is, the sighting of one object does not in any way impact on the sighting of another. These are assumptions that can be difficult to justify in many surveys of elusive animals or birds. Observations made with an angle gauge (Ståhl 1997, 1998) are possible too and for specific purposes they are efficient.

A characteristic of most line transect surveys is the nonconstant probability of detection of population elements. In surveys with human observers visual obstacles along the transect lines, the possible elusive nature of the elements of interest, and our limited field of vision combine to make the sightings imperfect. Many elements cannot be seen and others are simply missed. This phenomenon has to be taken into account in the estimation of a desired population statistic obtained from an imperfect transect line survey. Unbiased estimates of, say, population totals are only attainable if we know the probability of detecting a given population element given our location of observation. This assumption is rarely satisfied in practice. Often the probability of detection is some function of the distance between the elements and the survey line. Detection could be perfect up to a critical distance, after which it declines rapidly or it is only nearly perfect for elements on the survey line and then decreasing monotonically with distance (Ramsey and Harrison 2004). In repeat surveys within a single fixed region the surveyors may obtain solid information about the detection function (f), in others an estimate must be derived from the survey itself. When the detection function is derived from the survey data the estimated population statistics will only be approximately unbiased (Thompson 1992).

There are a plethora of line transect methods reflecting requirements and adaptations to the specific nature of the survey population and the environment in which it exists. Only the most common line transect survey design will be presented in this section. Other popular alternatives have been detailed by, for example, Ramsey et al. (1988).

The design we have chosen for detailing has a random selection of survey lines, with the survey lines selected with probability proportional to their length (L). The objective of the survey is to estimate a population total (N), viz., density N per hectare, of an attribute of interest, for example, the number of beetle-infested trees within a known fixed area representing a population or a natural stratum of host trees for the beetle. First a fixed baseline (B) for the reference and orientation of the transect lines is constructed. It is customary to let the baseline run parallel to one of the axes in an orthogonal reference coordinate system defining the outline of the population and all its elements. A number (n_{L}) of transect lines running orthogonal to the baseline and extending across the entire population are now chosen by simply generating n_{L} random locations within the population of interest. Since the number of points on a transect line is proportional to its length the procedure will automatically generate transect lines with probability proportional to their length. The baseline need not be a single line. For some populations with a very irregular outline or a very large spatial domain it is often advantageous to slice the populations by a system of parallel baselines. In that case the population is viewed as a series of disconnected slices each defined relative to their baseline. Selection of transect lines proceeds as for the case of a single baseline. Along the entire length of the ith survey line $\left(L_i, i = 1, \ldots, n_L\right)$ the surveyor records the shortest distance $\left(x_{ij}, j = 1, ..., n_i\right)$ from the survey line to each of the n_i sighted elements along the ith line. Alternatively, the surveyor records a sighting angle θ_{ij} and a distance r_{ij} and converts these two measurements to x_{ij} via $x_{ij} = r_{ij} \times \sin\theta_{ij}$. If sightings were perfect up to a distance of w_{100} with no sightings possible beyond this distance then the obvious (unbiased) density estimator for the ith survey line would be $D_i = n_i \times \left(2 \times w_{100} \times L_i\right)^{-1}$ and we would proceed to a length-weighted estimate of the population density and the sampling variance. Given that the area of the population is known, an estimate of the population totals is obtained by a simple multiplication of the density estimate and PA. However, we shall assume that detection is an unknown function of distance but no element at distance zero would be missed. With these assumptions the density estimator obtained from the ith transect line becomes

$$\hat{D}_i = \frac{n_i \times \hat{f}\left(0\right)}{2 \times L_i},$$

where $\hat{f}\left(0\right)$ is the estimated value of the detection function at a distance of zero. We detail the estimation of the detection function in the following. The density estimator is clearly a ratio of two random variables (n_i and L_i) and is

consequently biased. A jackknife estimator of the population density is therefore often preferable to a direct estimator (Efron 1982). A jackknife estimator reduces the first-order bias by taking the average of n_L leave-one-out estimates of the population density. Specifically we have

$$\hat{D}_{jk} = \frac{1}{n_L} \times \sum_{i=1}^{n_L} \hat{D}^{(i)} = \frac{1}{n_L} \times \sum_{i=1}^{n_L} \frac{\hat{f}(0) \sum_{i' \neq i} n_{i'}}{2 \times \sum_{i' \neq i} L_{i'}},$$

where $\hat{D}^{(i)}$ is a population density estimate obtained after excluding data from the ith transect line. The corresponding jackknife estimator of the sampling variance is

$$\hat{\text{var}}\left(\hat{D}_{jk}\right) = \frac{(n_L - 1)}{n_L} \times \sum_{i=1}^{n_L} \left(\hat{D}^{(i)} - \hat{D}_{jk}\right)^2.$$

Note, this variance estimator does not account for the uncertainty in the estimate of the detection function at distance zero nor the covariance between this estimate and the random variables n_i and L_i (Shenk et al. 1998). The omission is intentional since reliable estimates of these quantities require an intensive sampling of transect lines ($n_L > 30$) since we would need a separate estimate of $\hat{f}(0)$ for every transect line. Estimates derived from a small number of survey lines tend to be erratic. Confidence intervals for the population densities are obtained as outlined in Sect. 3.3.1.3.

The detection function can be estimated in a number of ways. A subjective but quick method derives $\hat{f}(0)$ from a histogram of the sighting distances in the survey. If there is a sharp drop in the number of sightings beyond a distance of, say, w_{100}, and one is willing to assume that no element was missed at shorter distances, $\hat{f}(0)$ would be estimated as $1/w_{100}$; conversely $\hat{f}(0)$ is estimated as the scaled height of the first class in a histogram with class intervals chosen in some suitable way (Wand 1997). Estimation by kernel smoothing (Izenman 1991) would convey attractive properties to $\hat{f}(0)$ but kernel-based estimation of the lower endpoint of a density function restricted to the domain of positive real numbers remain problematic. We opt for an estimation via the Fourier series method (pp. 67–70 in Silverman 1986). The Fourier series method of estimating $\hat{f}(0)$ is

$$\hat{f}(0) = \frac{1}{w^*} + \sum_k \hat{a}_k, k = 1, ..., n_{\text{obs}},$$

where w^* is the maximum distance at which an element can be sighted, n_{obs} is the total number of sightings in the n_L transect lines, and $\hat{a}_k$ are the Fourier coefficients given by

$$\hat{a}_k = \frac{2}{n_{\text{obs}} \times w^*} \sum_{i=1}^{n_L} \sum_{j=1}^{n_i} \cos\left(\frac{k \times \pi \times x_{ij}}{w^*}\right).$$

The number of Fourier coefficients $\hat{a}_k$ to include is determined by the first value of k for which

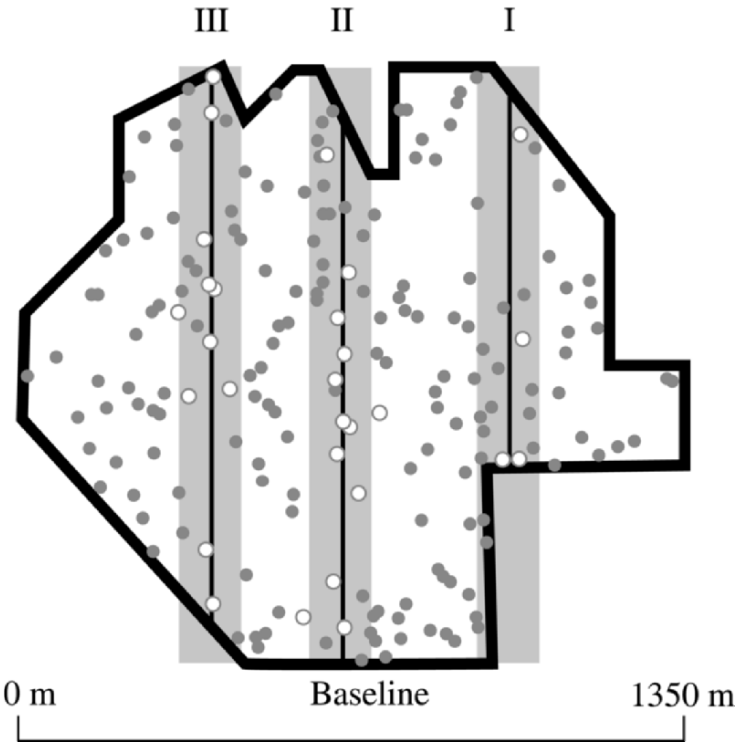


Fig. 3.22. Population outline with baseline and three transect survey lines. Population elements are indicated with *dots* and those observed from any of the three lines with *circles*. The *shaded area* around each survey line gives the true 95% detection interval

$$\frac{1}{W^*}\sqrt{\frac{2}{n_{\text{obs}}+1}} \geq \left|\hat{a}_k\right|$$

is true (Burnham et al. 1980). In practice, the number is rarely above 2.

A simple example will illustrate the estimation process. The population density of bark-beetle-infested trees within the 100.3-ha population domain shown in Fig. 3.22 is to be estimated by a transect survey with three random survey lines selected with probability to their length. The 1,350-m-long baseline and the three selected survey lines of length 756, 1,102, and 1,114 m and orthogonal to the baseline are indicated in the figure. The surveyor(s) move along the entire length of each survey length and record every infested tree (recognizable by resin exuding on the stem and possibly by a reddish needle discoloration) they can spot and then record the distance between the tree spotted and the survey line. Observations are not perfect: some trees will be missed and there is a natural (unknown) limit to the distance an infestation can be ascertained

Table 3.6. Number of sights on survey lines

j	Distance x_{ij} (m)		
	x_{1j}	x_{2j}	x_{3j}
1	17	1	3
2	15	25	3
3	16	8	37
4	22	15	11
5		70	4
6		30	2
7		82	67
8		1	17
9		14	6
10		10	4
11		12	42
12		2	
13		34	
Mean	17.5	23.4	17.8

from the survey line. The total number of sightings is 28. Table 3.6 gives the number of sightings on each survey line (4, 13, and 11) and the distances to each sighting from the survey line.

Sighted and nonsighted trees are indicated by different point signatures in Fig. 3.22. A histogram of the observed distances of elements from the survey lines is in Fig. 3.23. The sharp drop-off in the frequency of distances beyond

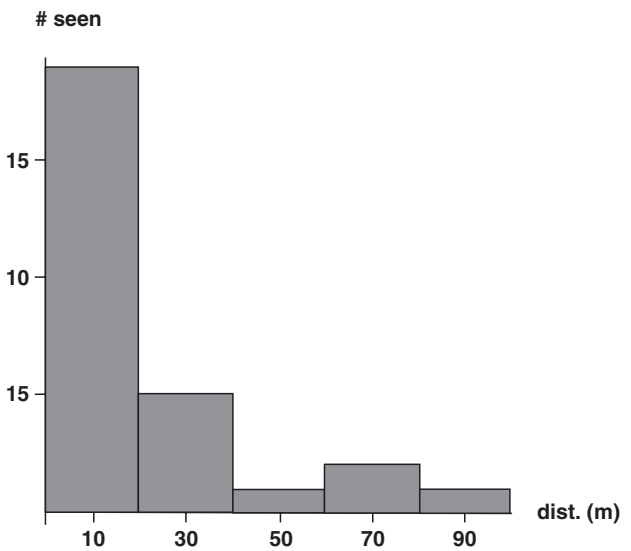


Fig. 3.23. Histogram of distances between seen objects and the nearest survey line

20 m is noticeable. We should expect w_{100} to be about 20 m. As an estimate of w^*, we took the 95% quantile of the recorded distances, which was 70.2 m. Recall that only sighted elements are recorded; hence, a 95% sample quantile corresponds to a higher quantile in the population of distances, in our case the 98% quantile. One should avoid the use of an observed maximum distance for w^* as it is a highly variable statistic; the three line-specific maxima bear this out. The number of Fourier series coefficients to use according to the previous rule is 1 and $\hat{a}_1 = 0.0153$, which gives us $\hat{f}(0) = 0.0295$ as the estimated fraction of perfect sightings. Delete-one jackknife estimates of density were accordingly 1.60, 1.18, and 1.35 with an average $\hat{D}_{jk} = 1.38$ and a relative standard error of 17.5%. A nominal 95% confidence interval runs from 0.34 to 2.42. The actual population contained 200 infested trees with a density of 1.81 trees per hectare.

A density estimate obtained from a single transect line does not have a design-based estimator of variance. Only an analytical estimate of the variance is possible, and only if the surveyor is willing to make assumptions about the distribution of elements and the arrangement of all possible survey lines inside the population. The correctness of these assumptions can be difficult to ascertain and the resulting estimates of variance can be quite poor. In the same vein, observed distances from either a single line or multiple lines can be used together with other predictors to generate spatial predictions of occurrence (Hedley and Buckland 2004). The quality of all estimators obtained from a line transect survey rests with the homogeneity of the detection function. If the probability of detection depends on more than distance then these factors must be incorporated into a generalized detection function (Marques and Buckland 2003; Ramsey and Harrison 2004). Finally, it is also critical that distance recordings are without errors. The impact of measurement errors can be serious and should be assessed whenever possible (Marques 2004). In a finite spatial domain of a population with a finite number of elements, any elements close to a population boundary will have a lower likelihood of detection than an element further away from the boundary. The reasons are the same as discussed for point sampling. The area of average detection is smaller for points close to the boundary because they can only be detected from locations inside the populations. The integral of all possible detection distances multiplied by the probability of detection is smaller owing to the restrictions imposed by nearby boundaries with nontrivial detection probability. When the area associated with points closer than w^* from a boundary only represents a small fraction of the total area, the boundary effect will be small. To gauge the potential of bias the surveyor can compute the reduction in the survey area due to boundaries. In our example, 95% of the recorded distances were within the gray bars in Fig. 322. The search area "lost" owing to boundary effects can be obtained by drawing six lines parallel to the baseline and going through the starting and end points of the three survey lines I, II, and III. Six triangles form between these six lines, the population boundary, and the outside of the shaded

95% detection limits. In the example, these areas account for 0.5% of the nominal (95%) search area in the absence of any boundary. It is, then, probable than the number of observed elements is biased downwards by this amount. In a larger survey with hundreds of observations it might be reasonable to add a matching fraction of the total count to the observations and assign an average distance to these "pseudo-observations."

3.7.4 Capture–Recapture Sampling

Capture–recapture sampling is primarily used for estimation of the population size of mobile population elements. Applications extend to estimation of the probability of detection in line transect surveys with a fixed survey width along the survey line (Borchers et al. 1998). In its simplest form a number of population elements (n_1) are captured at time 1 according to a chosen sample design and capture method, marked, and then released. At time 2 a new sample of n_2 elements is captured, of which n_{20} were unmarked and n_{21} were marked at time 1 $\left(n_2 = n_{20} + n_{21}\right)$. If it can be assumed that the total population size N is fixed during the time of the survey, that the first sample is representative of the population, that the n_1 marked elements distribute themselves uniformly across the population domain after the first capture, that the probability of catching a population element at time 2 is unaffected by the outcome of the first sample, and that the second sample is also representative of the population, then the minimum biased estimator of the population size is (Seber 1982)

$$\hat{N} = \left(n_1 + 1\right) \times \frac{\left(n_2 + 1\right)}{\left(n_{21} + 1\right)} - 1.$$

We prefer this estimator to Petersen's estimator of $n_1 \times n_2 \times n_{21}^{-1}$ (Seber 1982) as it is undefined for n_{21}=0. Implicit in this estimator is the assumption that the ratio of recaptured elements in the second sample extends to the population at large; an assumption that only holds if the two samples are truly representative of the population at large. Large sample sizes are needed before the assumptions can be fully justified. It should be noted that there is no unbiased estimator of N. Considerable effort has been spent on devising sample designs and capture methods that mitigate potential sources of bias (Seber 1986; Knight 2003; Wintle et al. 2004). An approximate unbiased estimator of the variance of this model-based estimator is

$$\hat{\text{var}}\left(\hat{N}\right) = \frac{\left(n_1 + 1\right)\left(n_2 + 1\right)\left(n_1 - n_{21}\right)\left(n_2 - n_{21}\right)}{\left(n_{21} + 1\right)^2\left(n_{21} + 2\right)}.$$

One of the most persistent problems of capture–recapture surveys is the potential for interactions between the population elements and the capture process. Models describing the effect of differential probabilities of capture at time 1

and time 2, birth and mortality processes, emigration, and immigration have been developed (Burnham et al. 1980; Norris and Pollock 1998; Efford 2004) in order to obtain model-based estimates of population size at the first, or second, survey time. Estimators of population sizes and their sampling variances for sampling on more than two occassions are given by, for example, Cormack (1993).

A maximum likelihood or a Bayesian estimation of N is possible if one is willing to make assumptions about the distribution of n_{12}, the only unknown random variable in the estimation problem. The distribution of n_{12} is usually assumed to be of hypergeometric, binomial, or Poisson form. In the binomial and Poisson model, N is a random variable, not fixed. A likelihood function can be associated with each of these models, which, in turn, would allow a likelihood-based estimation of n_{12}. In many cases some prior knowledge exists about what the distribution of n_{12} might be. Earlier surveys or subject knowledge could forward a prior distribution of n_{12} which would open up the possibility for a Bayesian estimation procedure (Poole 2002).

Shiver and Borders (1996, p. 333, example 11.4.1) illustrate a capture–recapture estimation problem with n_1=125, n_2=100, and n_{21}=44. The estimated population size using the previous estimator was 282 (rounded) with a standard error of 25 (rounded). Had we made the assumption of a hypergeometric distribution for n_{12} the estimated population size would have been 284 with a standard error of 29 (rounded). The variance of the maximum-likelihood estimate is obtained from

$$\hat{\text{var}}\left(\hat{N}_{\text{MLE}}\right) = \left(-\frac{\partial^2 \ell\left(n_{21} \mid n_1, N\right)}{\partial n_{21}^{\,2}}\bigg|_{N=\hat{N}} \right)^{-1} \times \left(\frac{n_2}{\hat{N}}\right)^2,$$

where $\ell\left(n_{21} \mid n_1, N\right)$ is the likelihood function of n_{21} and the last factor accounts for the scaling from the sample size (n_2) at time 2 to the estimated population size. Computation of the derivatives of the likelihood function is difficult and complex regardless of the model chosen. Software that can do symbolic calculations is needed for easy estimation.

3.7.5 Inverse Sampling

With modest sample sizes a low sample yield of marked elements (n_{21}) at time 2 puts the surveyor in a conundrum. Sample variation may simply have reduced n_{21} by chance but the ensuing estimator of N may be counterintuitive or apparently in error. Inverse sampling is a sample design in which the sample yield is fixed prior to sampling, which makes the sample size an unknown random variable (Panchapakesan et al. 1998; Cuzick 2001; Moore et al. 2003). The advantage is clear: a target yield is assured. The downside is equally clear:

there is no control over sample size and in practice the order of the sites to be visited until the target yield has been reached has to be random to avoid the sample being nonrepresentative of the population. For example, if the surveyor decides that n_{21}=44 then every single site to be visited before the target is reached has to be determined by a random draw of all possible sites. This may result in excessive travel times and cost and is only practical if the samples can arrive in random order at a fixed location of observation. While estimators under inverse sampling are generally close to or identical to the estimators under random sampling the estimated variances are typically much larger. The sampling distribution (of sample sizes) under inverse sampling at time 2 for recapture could be assumed to be of negative hypergeometric or negative binomial form (Johnson et al. 1992). To give an example of the variance inflation, suppose that the surveyor in the previous example had decided that a yield of 44 marked elements at time 2 would be desired and that this yield by chance was achieved after catching 100 elements. The maximum-likelihood estimate of N would again be 284 but now the standard error would be 44 (rounded), or almost 50% higher. Since there is no guarantee that the target yield can be obtained within the available time and with existing resources, and given that there is considerable risk of an inflated sampling variance, a cautionary approach to inverse sampling is prudent.

3.7.6 Double Sampling

Two independent surveys or a survey in combination with a registration system can be an efficient design for estimation of the total rare/elusive population. Let n_1 be the number of elements recorded during the first survey, n_2 be the number for the second survey, and n_{12} be the number identified in both surveys. Recorded objects must be identified clearly and uniquely in order to obtain n_{12}. Let N be the unknown population total that we wish to estimate. N is assumed constant from the onset of the first survey to the end of the second survey. For two independent surveys we expect to find $n_1 \times n_2 \times N^{-1}$ elements recorded in both surveys. Given the observed count n_{12}, we obtain a double-sampling estimate of N via

$$\hat{N}_{\mathrm{DS}} = \frac{n_1 \times n_2}{n_{12}}.$$

This estimator was first proposed by Chandra-Ssekar and Deming (1949). Owing to its simplicity it has found widespread applications in human surveys, and wildlife surveys. There is no need to have an exact estimate of the PA or for that matter an estimate at all in order to estimate the total, perhaps one of the main attractions of double sampling. No variance estimator has been forwarded for this double-sampling estimator of the total. When sampling is with

well-identified units, like line transects, survey lines, city blocks, or forest stands, then a jackknife estimator of variance and sampling error is recommended (Shao 1996). Note that the estimator is undefined for $n_{12}=0$ and is generally very unstable for small counts of n_{12}. It is our experience that n_{12} should be around 5% of $n_1 \times n_2$ before estimates with an acceptable accuracy (relative error less than 20%) can be expected. This means that the sampling intensity has to be rather high in both surveys; if n_1/N or n_2/N sinks below 10% the chance that $n_{12}=0$ is nontrivial. A more specific assessment would require assumptions about both survey design and population statistics, such as total and distribution across the spatial domain.

Double sampling in forest inventory could be an option for the estimation of, for example, the total number of trees of a rare species, the number of stems logged at a logging site, the number of trees fallen owing to windthrow, or the number of diseased trees. Counting could be done along random survey lines with markings of all observed elements falling on the line(s) or in close proximity to the line(s) or it can be done on remotely sensed images. We shall illustrate the double-sampling design for the estimation of the number of windthrown trees in the same forest we used for demonstrating line transect sampling, capture–recapture, and sampling with PPS.

A severe storm felled 1,688 trees in the 110.2-ha forest. The damage was mostly concentrated in 11 areas (12% of total) of size 0.1–4.2 ha (average 1.2 ha) but scattered fall downs were observed throughout the forest. The number of trees downed in the hardest-hit areas ranged from 41 to 59 per hectare with a mean of 50 stems per hectare. Tall trees were predominantly hit by the storm. The stem length of the fallen trees was 51 m, with a standard deviation of about 5 m. The surveyor decides to assess the damage by laying out two independent surveys, each with 30 random survey lines, random with respect to location and length. The orientation was random within a limited range of angles. The average length of a survey line is approximately 275 m but individual lines range from 30 to 890 m. A map of the fallen trees and the two sets of survey lines is shown in Fig. 3.24.

In the first survey 266 fallen stems crossed the survey lines (n_1), while 200 crossed the lines of the second survey (n_2). A total of 41 stems were common to both surveys (n_{12}). From this we get an estimated total of 1,298 fallen trees. A jackknife estimator of the total was 1,309, which indicates a bias of 11 (1%) in the double-sampling estimator. The jackknife estimator of the standard error was 122. To obtain the jackknife estimators we deleted one survey line from the first survey and one survey line from the second survey number; hence, 900 delete-one estimates were obtained. The distribution of these delete-one estimates of the number of fallen trees is shown in the histogram in Fig. 3.25. Notice the skewed distribution and the appearance of a mixture of two distributions arising from the spatial heterogeneity of the intensity of

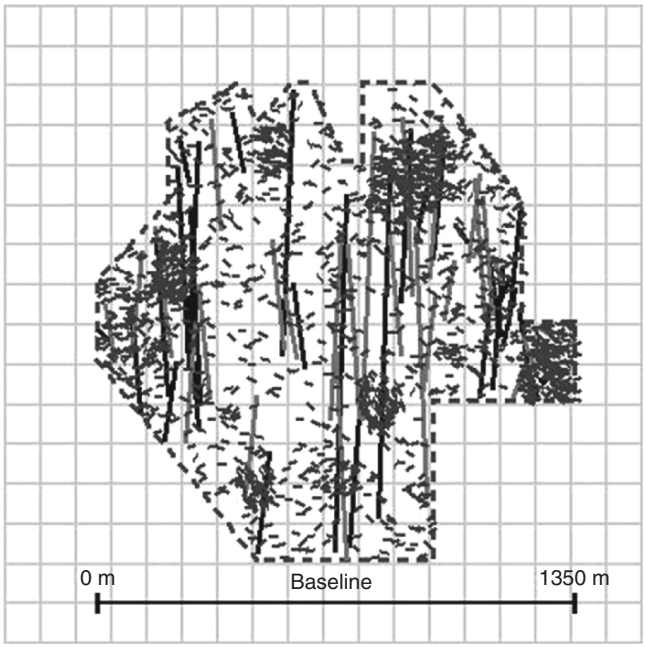


Fig. 3.24. Map of fallen trees and survey lines in the first survey (*black*) and in the second independent survey (*gray*). The grid spacing is 100 m × 100 m

windthrow across the forest. The fact that the nominal 95% confidence interval for the sample-based estimator based on the assumption of a normally distributed sample estimate does not include the true number is a direct consequence of this skewed distribution. A confidence interval with a coverage closer to the nominal value should, therefore, be obtained from the quantiles of the bootstrap distribution of sample estimates (Shao 1996).

3.7.7 Composite Sampling

In sampling for rare/elusive population elements the time and cost to identify the presence/absence or to quantify the attribute of interest in a sample unit can be very costly. For a rare/elusive element most sample units will have a value of zero but will still carry the full cost of analysis. Soil sampling for rare contaminants, sampling containers of wood chips for the presence of a rare staining fungi or nematode, or landscapes for rare deforestation events are but a few examples with relevance to forest inventory. As the name implies, in composite sampling several sample units are joined into a single composite unit.

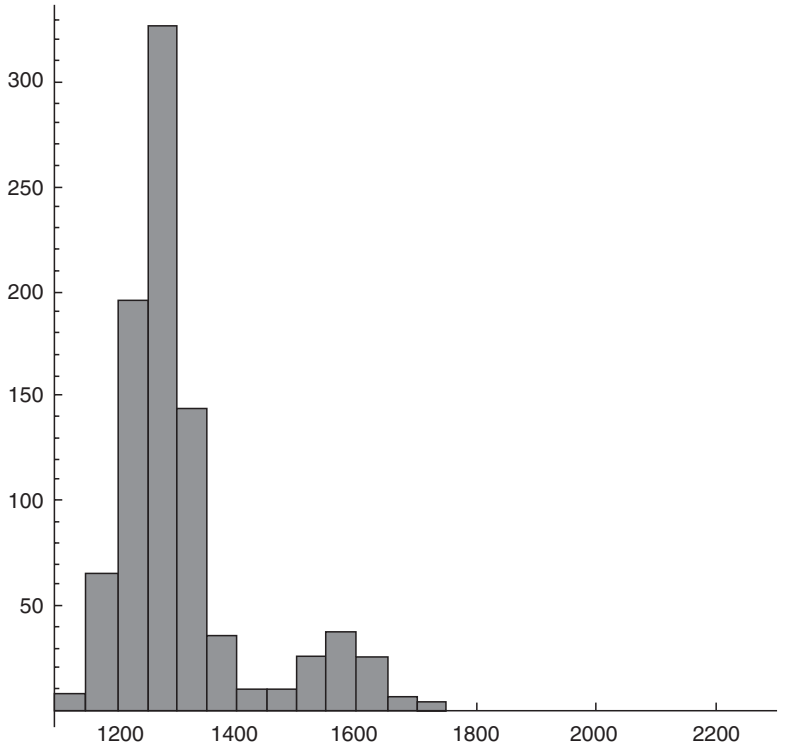


Fig. 3.25. Histogram of jackknife (delete-one) estimates of the number of fallen trees

The idea behind composite sampling is simple enough: the screening for the presence/absence of an elusive element is concentrated on fewer composite units. The only sample units to be examined individually are those for which the composite sample came up positive for the presence of the rare element. Lancaster and Keller-McNulty (1998) have reviewed the composite sampling method and they provide a succinct overview of the pros and cons of this method. Composite sampling is a process that involves defining and optimizing the compositing design, the measurement process, and the data analysis process. The observed composite response y for composite unit i can be represented by

$$y_i = f\left(x_{i1}, x_{i2}, \ldots, x_{in_i}; \mathrm{c}\right) + e_i, i = 1, \ldots, n,$$

where x_{ij} is the attribute value of the jth sample units in the ith composite unit, $f\left(\cdot;\cdot\right)$ is a function of the physical process of compositing, **c** is a set of weights that depend on assumptions made about the physical process, and e represents the measurement error. The utility of this expression may not be immediately

clear or useful unless we know the function *f*, the vector **c**, and the measurement error. When the compositing process model and the errors are not fully known the inference about *y* becomes model-based and the uncertainty regarding model parameters must be accounted for in the estimators of sampling variance. In its simplest form, when the statistical objective is to estimate a population mean, the physical process is presented as

$$f\left(x_{i1},\ldots,x_{in_i};\mathrm{c}\right)=\sum_{j=1}^{n_i}c_{ij}\,x_{ij}.$$

If the sample units enter with equal amount then the expected value of *y* is the population mean with a variance of $\sigma_{\mathrm{x}}^2/n+\sigma_{\mathrm{e}}^2$. Compare this with the variance of $\sigma_{\mathrm{x}}^2/n+\sigma_{\mathrm{e}}^2/n$ that one would expect if the *n* measurements were done on *n* sample units. Hence, composite sampling is a trade-off: cost savings are achieved at the expense of precision and information. The trade-off will have to be weighed carefully in each case. More elaborate schemes are possible: sample units can themselves be sampled before they are combined into composite units and the composite units can, in turn, also be sampled. Finally, measurement units may be a small fraction of a composite unit. The process of mixing and subsampling is captured by the vector **c** and the error term. Lancaster and Keller-McNulty give a good example of how the sampling designs can be employed to illicit estimates associated with population features, such as, for example, row and column effects in spatial sampling. Estimation of prevalence is also possible from composite sampling but model assumptions that need to be verified are needed. Commonly one assumes that *x* is a Bernoulli random variable with probability π_x of taking the value of 1. If *r* sample units are pooled equally into a composite sample then the probability that y_i is 1 becomes $1-(1-\pi_x)^r$. A maximum-likelihood estimator of prevalence is

$$\hat{\pi}_x=1-\left(1-\frac{\sum_i y_i}{n}\right)^{\frac{1}{2}}.$$

A note of caution is in place: the estimator is not unbiased and it can be quite imprecise, especially for high values of *r*.

Classification of pixels sampled from a remotely sensed image is akin to composite sampling when the size of the pixel is a multiple of the sampling unit applied in forest inventory. What is observed is a composite of features in several individual inventory units. Attempts to obtain estimates at the scale of sample units by "unmixing" (Oleson et al. 1995; Bosdogianni et al. 1997; Grandell et al. 1998; Mertens 2003; Vikhamar and Solberg 2003) are in essence attempts to solve for *x* given *y* and a model for **c**. The problem is generally underdetermined (more variables than equations) but if one is willing to specify *f* and assume **c** to be invariant then one can use ordinary least-squares or mixed linear models for the estimation of **c** and ultimately estimators of *x*.

3.8 Small-Area Estimation

A forest inventory is designed to provide estimates of attribute values of interest for the entire population or for a set of strata identified during the design phase. After completion of an inventory it often happens that attribute estimates or an update of estimates for one or several small geographic areas of the sampling frame of the inventory is needed. Estimators pertaining to the forest in a particular county, district, or any other zone of interest to someone are therefore needed on a routine basis. Related postsampling estimation problems have surfaced in most large-scale surveys and spawned a search for effective, design-consistent estimators applicable to domains within a population and small geographically defined areas (Särndal et al. 1992; Rao 2003). A common feature in all small-area estimation problems is the small number of samples taken within the small area. A direct estimation based on only the samples taken inside the small area would in most cases provide estimates with low precision and that are possibly biased. The survey context, the particular features of the small area or domain, and the availability of auxiliary information determine in each case the most promising approach to estimation. A rich and diverse collection of estimators have been tailored to a wide spectrum of small-area estimation problems. The majority are model-based or at least model-assisted. The data for a small-area or domain are assumed to adhere to a model that we wish to estimate. Optimality of estimates in terms of minimum bias and minimum variance is the ideal that is pursued but it is rarely achieved. Most estimators are in some sense the "best possible" given the estimation problem posed. One of the first published forestry applications was given by Green et al. (1987). Timber volumes per hectare and by county were to be estimated from a regionwide inventory. They assumed that county-specific sample-based estimates (means) of timber volumes per hectare were estimates of the sum of a random county-specific effect and a random "error." Improved, in terms of mean squared error, county-specific estimates were then obtained as a weighted average of the county-specific sample means and a weighted mean of all counties. Related estimation problems are recurrent in forestry and with the increase in the use of remotely sensed data as auxiliary information we have greatly expanded our options for effective small-area estimation (Kangas 1996; Lappi 2001). Only a few of the most commonly applied estimators will be presented here. Rao (2003) provides a recent summary of small-area estimators and Pfeffermann (2002) offers a review of current trends, unresolved issues, and the future direction of small-area estimation research.

3.8.1 Direct Small-Area Estimators

Direct small-area estimation means that only the samples collected in the small area (U_i) are used for estimation purposes. The sample size n_i in U_i is a random variable with a possible value of zero. As long as the probability of n_i=0 remains low and n_i is close to its expected value, the estimators employed for estimation of population level attributes apply (Särndal et al. 1992; Rao 2003). Often, however, there is a risk that n_i=0 is nontrivial and this must be considered in estimating a small-area sampling variance.

Under SRS in a population occupying an area A, with n fixed-area plots each with an area A_{plot} the direct small-area estimator for U_i say a total Y, is

$$\hat{Y}_{i\,(\text{dir}\,|\,\text{SRS})} = \frac{A}{n \times A_{\text{plot}}} \sum\nolimits_{j \in U_i} Y_j,$$

with an approximate variance of

$$\hat{\text{var}}\left(\hat{Y}_{i\,(\text{dir}\,|\,\text{SRS})}\right) = \left(\frac{A_i \times n}{A \times n_i}\right)^2 \left(\frac{A}{n \times A_{\text{plot}}}\right)^2 \left(1 - n \times \frac{A_{\text{plot}}}{A}\right)\left(\frac{n_i - 1}{n - 1}\right) \times \text{var}\left(Y_j \,|\, j \in U_i\right),$$

where A_i is the area of U_i and Y_j is the plot total in the jth plot in U_i. Direct variance estimation is obviously only possible when $n_i \geq 2$. When the area of U_i is not known or only known with error, the added uncertainty stemming from either predicting A_i from the ratio n_i/n or the error in A_i must be factored in. In that case a variance approximation based on a Taylor series approximation would be appropriate.

Direct estimates can be improved if the small area can be stratified into G groups based on an attribute value closely related to the attribute of interest. The sample size in each of the G groups must be larger than or equal to 2. In that case a poststratified direct estimate for the small area $\hat{Y}_{i\text{dir}\,(\text{SRS}\,|\,\text{poststrat})}$ will have less variance than a simple direct estimate. $\hat{Y}_{i\text{dir}\,(\text{SRS}\,|\,\text{poststrat})}$ is obtained as per stratified random sampling, but the approximate variance of $\hat{Y}_{i\text{dir}\,(\text{SRS}\,|\,\text{poststrat})}$ is

$$\hat{\text{var}}\left(\hat{Y}_{i\text{dir}\,(\text{SRS}\,|\,\text{poststrat})}\right) = \sum_{g=1}^{G} \left(\frac{n}{n-1} \times \frac{n_{ig} - 1}{n_{ig}}\right) \times \left(\frac{A_{ig}}{n_{ig} \times A_{\text{plot}}}\right)^2 \times \left(\frac{1}{n_{ig}} - \frac{n}{A \times n_{ig}}\right) \times \text{var}\left(Y_{ig}\right),$$

where n_{ig} is the sample size in the intersection of the gth group with U_i and Y_{ig} is the plot total of y in group g plots in U_i.

3.8.2 Synthetic Small-Area Estimators

A small-area estimator is called synthetic if it is obtained from a larger area, covering several small areas, under the assumption that the large population is homogenous with respect to the attributes of interest. In its simplest form the estimator of the mean value of an attribute in the *i*th small area becomes $\hat{\bar{Y}}_{i\,(\mathrm{syn})} = \hat{\bar{Y}}$, where $\hat{\bar{Y}}$ is the sample-based estimate of the population mean. Small-area totals are obtained by multiplying the mean by the area of the small area (assumed known without error). A synthetic estimator benefits from the (usually) low MSE of $\hat{\bar{Y}}$, but suffers from a potentially serious bias if the small area differs from the area of the rest of the population. Despite the obvious potential problem in applying a population mean to a small area, its precision makes it harder to ignore. Conversely an estimate based solely on data from the small area $\left(\hat{Y}_i\right)$ may appeal on the grounds of bias but not in terms of precision. As we shall see, a compromise in the form of a weighted average of a small-area estimate and global information can strike a good balance.

Auxiliary information in the form of a vector $\mathbf{x}$ related to y through the population model $y = \mathbf{x}' \cdot \boldsymbol{\beta}$ may be available in the form of known totals $\mathbf{X}_i$ for the small area. A synthetic regression estimator of the total is then $\hat{Y}_{i\mathrm{syn}_{\mathrm{reg}}} = \mathbf{X}'_i \cdot \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is an estimate of the population regression coefficients (c.f. Sect. 3.3.5.1). The bias of the regression estimator will be small if the small-area regression coefficients $\beta_i \approx \beta$, an assumption that can be examined more closely in a separate assessment of the model (Ronchetti et al. 1997; Zhang and Davidian 2001). A special case of the synthetic regression estimator is the synthetic ratio estimator $\hat{Y}_{i\,(\mathrm{syn\,|\,ratio})} = X_i \times \hat{R}$, where X_i is the total of x in the *i*th small area and $\hat{R}$ is the ratio of the estimated population totals of y and x. Since the synthetic estimators can be biased an estimate of their MSE is needed (variance plus squared bias) to gauge precision. Design-consistent variance estimators of the synthetic estimators, $\hat{\mathrm{var}}\left(\hat{Y}_{i\,(\mathrm{syn})}\right)$, are obtained as per the design employed, but a reliable estimate of bias is harder to obtain. A common approach to estimate the MSE of a synthetic estimate is to obtain synthetic and direct sample-based estimates for a group of m "similar" small areas and then compute

$$\widehat{\mathrm{MSE}}\left(\hat{Y}_{i\,(\mathrm{syn})}\right) = \hat{\mathrm{var}}\left(\hat{Y}_{i\,(\mathrm{syn})}\right) + \mathrm{bias}^2\left(\hat{Y}_{i\,(\mathrm{syn})}\right),$$

$$\mathrm{bias}^2\left(\hat{Y}_{i\,(\mathrm{syn})}\right) = \overline{\mathrm{MSE}}_m\left(\hat{Y}_{i\,(\mathrm{syn})}\right) - \frac{1}{m}\sum_{j=1}^{m} \hat{\mathrm{var}}\left(\hat{Y}_{j\,(\mathrm{syn})}\right),$$

and

$$\overline{\mathrm{MSE}}_m\left(\hat{Y}_{i\,(\mathrm{syn})}\right) \approx \frac{1}{m}\left(\sum_{j=1}^{m}\left(\hat{Y}_{j\,(\mathrm{syn})} - \hat{Y}_j\right)^2 - \sum_{j=1}^{m} \hat{\mathrm{var}}\left(\hat{Y}_j\right)\right).$$

Estimators applicable to small-area means are obtained in a similar way after the appropriate scaling. Provided that sampling errors dominate bias the previous estimator of MSE is robust and asymptotically design-consistent for $m, n_j (j = 1,\ldots, m) \to \infty$. Variance estimators based on resampling or leave-one-out jackknifing are often preferred to design-based estimators of variance. Synthetic estimators can, just like direct estimators, frequently be improved by poststratification into G groups as outlined for the direct estimators.

3.8.3 Composite Small-Area Estimators

We saw that a synthetic estimate could be seriously biased if the small area was distinctly different from the general population and that a local area estimate could be very imprecise. As a compromise the composite estimator provides a weighted average of two available estimators for the ith small area, say $\hat{Y}_{i1}$ and $\hat{Y}_{i2}$ of totals

$$\hat{Y}_{i\,(\mathrm{comp})} = \phi_i \times \hat{Y}_{i1} + \left(1 - \phi_i\right) \times \hat{Y}_{i2},$$

where $0 \le \phi \le 1$. Many small-area estimators have the composite form. Composite estimators of means are obtained in a similar way. The MSE of the composite estimator of a small-area total is given by

$$\widehat{\mathrm{MSE}}\left(\hat{Y}_{i\,(\mathrm{comp})}\right) = \phi_i^2 \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i1}\right) + \left(1 - \phi\right)^2 \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i2}\right) + 2\phi_i\left(1 - \phi_i\right) \mathrm{E}_p\left(\hat{Y}_{i1} - Y_i, \hat{Y}_{i2} - Y_i\right),$$

where $E_p\left(\hat{Y}_{i1} - Y_i, \hat{Y}_{i2} - Y_i\right)$ is the expected mean-square cross-product of the two estimators taken over all possible samples under the employed design (p). By choosing the weights that minimize the MSE one obtains

$$\widehat{\mathrm{MSE}}_{\mathrm{opt}}\left(\hat{Y}_{i\,(\mathrm{comp})}\right) = \frac{\widehat{\mathrm{MSE}}\left(\hat{Y}_{i1}\right) \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i2}\right) - \widehat{E}_p^{\,2}\left(\hat{Y}_{i1} - Y_i, \hat{Y}_{i2} - Y_i\right)}{\widehat{\mathrm{MSE}}\left(\hat{Y}_{i1}\right) \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i2}\right) - 2\,\widehat{E}_p\left(\hat{Y}_{i1} - Y_i, \hat{Y}_{i2} - Y_i\right)}$$

The difficulty with this estimator is that only approximate estimates of the MSE of cross-products between $\hat{Y}_{i1}$ and $\hat{Y}_{i2}$ can be obtained by application of some form of data resampling consistent with the design (Efron and Tibshirani 1993; Shao 1996; Lahiri 2003; Shen et al. 2004). If one is willing to assume that the MSE of the cross-product is negligible compared with the MSEs of Y_{i1} and Y_{i2}, the approximately optimal weight becomes

$$\varphi^*_{\mathrm{opt}} \approx \mathrm{MSE}\left(Y_{i2}\right) \times \left[\mathrm{MSE}\left(Y_{i1}\right) + \mathrm{MSE}\left(Y_{i2}\right)\right]^{-1}$$

and

$$\widehat{\mathrm{MSE}}_{\mathrm{opt}}\left(\hat{Y}_{i\,(\mathrm{comp})}\right) \approx \frac{\widehat{\mathrm{MSE}}\left(\hat{Y}_{i1}\right) \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i2}\right)}{\widehat{\mathrm{MSE}}\left(\hat{Y}_{i1}\right) \times \widehat{\mathrm{MSE}}\left(\hat{Y}_{i2}\right)} = \varphi^{*}_{\mathrm{opt}} \times \widehat{\mathrm{MSE}}\left(Y_{i1}\right)$$

from which we see that the MSE of the composite estimate is less than the smallest of the component MSEs. The maximum reduction is 50% below the smallest value which is achieved when the components receive equal weights. Under a SRS design in a homogenous population the composite estimator that combines an estimate based on samples inside the *i*th small area with one for the population at large becomes

$$\hat{Y}_{i\,(\mathrm{comp}\,|\,\mathrm{SRS})} = \frac{n_i}{N_i} \times \hat{Y}_i + \left(1 - \frac{n_i}{N_i}\right) \times \hat{Y},$$

which is also the best linear unbiased prediction (BLUP). Under similar circumstances and assuming that the variance of y is proportional to x the composite ratio estimator becomes

$$\hat{Y}_{i\,(\mathrm{comp}\,|\,\mathrm{ratio})} = \frac{X_i}{X} \times \hat{Y}_{i\mathrm{ratio}} + \left(1 - \frac{X_i}{X}\right) \times \hat{Y}_{\mathrm{ratio}}.$$

Multivariate composite estimators are obtained by simple extensions of the univariate estimators (Gregoire and Walters 1988).

3.8.4 Model-Based Small-Area Estimation

A model describing an attribute value of a population element as a linear combination of fixed large-scale effects and random local effects offers the most general and flexible approach for small-area estimation. Fixed effects are constant for all population elements, while the local effects are specific to a small area. The population or a large part of the population is viewed as an ensemble of several small areas and estimation is done for all members of the ensemble simultaneously. A model-based simultaneous small-area estimation approach offers the advantage that the estimate for a specific small area can be improved by "borrowing" information from either the entire ensemble of small areas or a specific subset of small areas with certain attributes in common. The model and the associated model assumptions detail the communality of attribute values between population elements within a single small area and between population elements in different small areas. A set of nested models is often necessary to succinctly describe the relationship between observed sample values in various parts of the population. Our ability to obtain robust design-consistent and asymptotically unbiased estimates of local random effects have improved dramatically over the last 2 decades and continue to do

so (Pfeffermann 2002). Two versatile yet simple models, the area-level linear mixed model and the element linear mixed model, will be given as examples. Checking of model adequacy, model fit, and model assumptions is incumbent upon the analyst whenever a model-based inference is deemed appropriate (Ritz 2004).

In the area-level model we assume that a small-area effect θ_i, $i = 1, \ldots, m$ is a known function $g\left(\bar{Y}_i\right)$ of a small-area attribute value, say the mean $\bar{Y}_i$, and furthermore is related to p area-specific auxiliary values $\boldsymbol{X}_i$ through the linear model

$$\theta_i = X'_i\beta + b_i \times v_i,$$

where the b_i are known positive constants and β is the $p \times 1$ vector of population-specific regression coefficients, and the v_i are area-specific random effects assumed to be independent and identically distributed with an expected value of zero and a variance σ_v^2. The function g can be the identity function, a linear function, or a nonlinear function, while the constants b_i are introduced to allow for heterogeneity in the variance of random effects. Note that the expectations of the random effects are with respect to the model, an important issue since it can be difficult to justify for areas for which n_i is small (less than 10).

We are interested in obtaining the BLUP of θ_i, which means that we seek a design-consistent minimum-variance estimator of $\boldsymbol{\beta}$ and a BLUP of v_i. Preliminary estimates of θ_i can be obtained directly from the sample as $\hat{\theta}_i = g\left(\hat{\bar{Y}}_i\right)$ but they are not the BLUP. We can write our direct sample-based estimates as an observational equation as

$$\hat{\theta}_i = \theta_i + e_i \text{ with } E_p\left(e_i \middle| \theta_i\right) = 0 \text{ and var}\left(e_i \middle| \theta_i\right) = \psi_i \text{ known}$$

Combining the model with the observational equation, we get

$$\theta_i = X'_i\boldsymbol{\beta} + b_i \times v_i + e_i,$$

which is a special case of a linear mixed model. The mix of sampling errors (e_i) and random model effects (v_i) makes the model rather unique and introduces inferential complexity. Especially, the assumption of known area-specific sampling variances may be viewed as restrictive, and typically a direct estimate or some smoothed estimate $\hat{\psi}_i$ is used in place of ψ_i.

Since we must rely on estimated variance components our estimators are no longer the BLUP but the empirical best linear unbiased prediction (EBLUP) (Wolter 1985). The EBLUP of θ_i is

$$\hat{\theta}_{i\,\text{EBLUP}} = \hat{\gamma}_i \times \hat{\theta}_i + \left(1 - \hat{\gamma}_i\right) X'_i\,\tilde{\boldsymbol{\beta}},$$

where

$$\hat{\gamma}_i = \hat{\sigma}_v^2 \times b_i^2 \times \left(\psi_i + \hat{\sigma}_v^2 \times b_i^2\right).$$

We recognize in $\hat{\theta}_{i\,\mathrm{EBLUP}}$ a composite estimator of a direct design-consistent estimate $\left(\hat{\theta}_i\right)$ and a synthetic estimate $\left(X'_i\tilde{\boldsymbol{\beta}}\right)$ for the ith small area with weights $\left(\gamma_i\right)$ determined by the strength of the among-area variation $\left(\sigma_v^2\right)$ relative to that of the total random variation $\left(\psi_i+\sigma_v^2\right)$. More weight is given to a direct local estimate when the data point to strong local effects and vice versa, an intuitively appealing attribute. Only area-level auxiliary variables ($\mathbf{X}_i$) are used for the estimation, which makes the estimate $\hat{\theta}_{i\,EBLUP}$ valid for any statistically valid sampling design. When $\theta_i = \boldsymbol{X}'_i\tilde{\boldsymbol{\beta}} + b_i \times v_i$ holds, the average bias will be zero. Estimators of $\boldsymbol{\beta}$ depend on an available estimate of σ_v^2 and vice versa; therefore, an iterative estimation process is needed. A current estimate of $\boldsymbol{\beta}$ is obtained from a current estimate of σ_v^2 and so on until convergence is achieved. Current estimates of $\boldsymbol{\beta}$ and σ_v^2 are

$$\tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^{m} \boldsymbol{X}_i\hat{\theta}_i \times \left(\psi_i + \hat{\sigma}_{\mathrm{v}}^2\right)^{-1}\right]\left[\sum_{i=1}^{m} \boldsymbol{X}_i\boldsymbol{X}'_i \times \left(\psi_i + \hat{\sigma}_{\mathrm{v}}^2\right)^{-1}\right]^{-1}$$

and

$$\hat{\sigma}_v^{\,2} = \hat{\sigma}_v^{\,2}\left|\sum_{i=1}^{m}(\hat{\theta}_i - X'_i\tilde{\boldsymbol{\beta}})^2 \times \left(\psi_i + \hat{\sigma}_{\mathrm{v}}^2\right)^{-1} - m + p \cong 0,\right.$$

where $\hat{\sigma}_v^2$ is a method of moments estimator of σ_v^2. Alternative estimators each relying on a set of specific assumptions are abound. The specifics of the data at hand and the experience of the analyst decide the choice.

Estimators of the MSE of $\hat{\theta}_{i\,\mathrm{EBLUP}}$ are approximate only since we rely on model-based estimates of model parameters and a sample-based estimate of error variances. The estimates are generally also biased. It is important to note that the estimation of MSEs should be tailored to the estimation procedure used for the fixed and random effects (Rao 2003). A slightly conservative MSE estimator that is valid for the previous estimate is

$$\widehat{\mathrm{MSE}}\left(\hat{\theta}_{i\,\mathrm{EBLUP}}\right) = \hat{g}_{1i}\left(\hat{\sigma}_v^2\right) + \hat{g}_{2i}\left(\hat{\sigma}_v^2\right) + 2\hat{g}_{3i}\left(\hat{\sigma}_v^2\right),$$

where

$$\hat{g}_{1i}\left(\hat{\sigma}_v^{\,2}\right) = \hat{\gamma}_i\psi_i,$$

$$\hat{g}_{2i}\left(\hat{\sigma}_{\mathrm{v}}^{\,2}\right) = \left(1-\hat{\gamma}_i\right)^2 X'_i X_i\left[\sum_{i=1}^{m} \boldsymbol{X}'_i\boldsymbol{X}_i \times \left(b_i\hat{\sigma}_{\mathrm{v}}^{\,2} + \psi_i\right)^{-1}\right],$$

and

$$\hat{g}_{3i}\left(\hat{\sigma}_{\mathrm{v}}^{\,2}\right) = \left[b_i^4\psi_i^2 \times \left(b_i\hat{\sigma}_{\mathrm{v}}^{\,2} + \psi_i\right)^{-4}\right]\left(\hat{\theta}_i - \boldsymbol{X}'_i\,\tilde{\boldsymbol{\beta}}^2 \times \hat{\mathrm{v}}\mathrm{ar}\left(\hat{\sigma}_{\mathrm{v}}^2\right)\right),$$

where $\hat{\mathrm{v}}\mathrm{ar}\left(\hat{\sigma}_v^{\,2}\right)$ is an estimate of the variance of an estimated variance component. A jackknife estimate of $\hat{\mathrm{v}}\mathrm{ar}\left(\hat{\sigma}_v^{\,2}\right)$ can be obtained by repeating the previous estimation procedures m times, each time with one different small area excluded from the analysis. Alternatively one can approximate this variance by

$2\hat{\sigma}_v^4 \times (n - m - p)^{-1}$. Datta et al. (1991) detailed a multivariate extension of the area-level linear mixed model.

Our second small-area model is the element linear mixed model. In this model the attribute value (y) of the jth individual population elements within the ith small area is modeled as a linear combination of p fixed $(\boldsymbol{x}'_{ij}\boldsymbol{\beta})$ effects known for every element in i and two random effects $(v_i + b_{ij}e_{ij})$

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}' + v_i + b_{ij}e_{ij},\ j = 1,\ldots,n_i, i = 1,\ldots,m,$$

where b_{ij} are known constants and e_{ij} is assumed to be a random variable with an expected value of 0 with respect to the model and a variance of σ_v^2. Again, $\boldsymbol{\beta}$ is a vector of population-level design-consistent regression coefficients. For estimation purposes it is often assumed that the distribution of the random variables is normal. We assume that a sample of size n_i has been taken from the N_i elements in the ith small area $\left(n = \sum_{i=1}^{m} n_i\right)$ and that this sample is consistent with the model. We wish to estimate, say, $\bar{Y}_i$, the mean of y in the ith small area. SRS from the ith small area or a sample selection based on $\mathbf{x}_{ij}$ both satisfy an appeal to validity of the generic model (Rao 2003) Under the element linear model the EBLUP estimator of the ith small-area mean can be written as a composite estimator of the survey regression estimator and the regression synthetic estimator:

$$\hat{\bar{Y}}_{i\,\mathrm{EBLUP}} = \hat{\gamma}_i + \left[\hat{\bar{Y}}_i + \left(\bar{\boldsymbol{X}}_i - \hat{\bar{\boldsymbol{X}}}_i\right)'\tilde{\boldsymbol{\beta}}\right] + (1 - \hat{\gamma}_i) \times \bar{\mathrm{X}}'_i\tilde{\boldsymbol{\beta}},$$

where $\hat{\bar{Y}}_i$ is the direct small-area estimate of $\bar{Y}_i$, $\bar{\boldsymbol{X}}_i$ is the $p \times 1$ vector of known small-area means of the auxiliary variables, $\hat{\bar{\boldsymbol{X}}}_i$ is the small-area sample estimate of $\bar{\boldsymbol{X}}_i$, and γ_i is the weight given to the survey regression estimator. Note, when the constants b_{ij} are not all 1 the sample means become

$$\hat{\bar{Y}}_i = \frac{\sum_{j=1}^{n_i} b_{ij}^{-2} y_{ij}}{\sum_{j=1}^{n_i} b_{ij}^{-2}},\ \hat{\bar{\mathrm{X}}}_i = \frac{\sum_{j=1}^{n_i} b_{ij}^{-2} x_{ij}}{\sum_{j=1}^{n_i} b_{ij}^{-2}}.$$

The best linear unbiased estimator of the population-level regression coefficients is

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} b_{ij}^{-2}\mathrm{x}_{ij}\mathrm{x}'_{ij} - \hat{\gamma}_i b_{i\cdot}^{-2}\hat{\bar{\mathrm{X}}}_i\hat{\bar{\mathrm{X}}}'_i\right)^{-1}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} b_{ij}^{-2}\mathrm{x}_{ij}y_{ij} - \hat{\gamma}_i b_{i\cdot}^{-2}\hat{\bar{\mathrm{X}}}_i\hat{\bar{Y}}_i\right),$$

where $\mathrm{b}_{i\cdot}$ is the sample sum of fixed variance constants b_{ij} for the ith small area. The weight given to the direct survey regression estimate is

$$\hat{\gamma}_{\mathrm{i}} = \hat{\sigma}_{\mathrm{v}}^2 \times \left(\hat{\sigma}_{\mathrm{v}}^2 + \hat{\sigma}_{\mathrm{e}}^2 \Big/ \sum_{j=1}^{n_i} b_{ij}^{-2}\right).$$

Provided we have large small-area sample sizes and $b_{ij} \equiv 1 \forall \{i,j\}$, the survey regression estimator is approximately design-unbiased but the synthetic regression estimator $\overline{\mathbf{X}}'_i \tilde{\boldsymbol{\beta}}$ may be a biased for $\hat{\overline{Y}}_i$. Estimation of the variance components σ_v^2 and σ_e^2 can proceed in different directions depending on the assumptions made and the preferences of the analyst. Maximum-likelihood and restricted-maximum-likelihood estimation requires assumptions about the distribution of the random effect. If warranted, these methods lead to more efficient estimates but only if the distributional assumptions actually hold. They also provide a generic framework for estimation regardless of the values chosen for the b_{ij}. As done for the area-level model we shall present the method of moment estimation procedures under the assumption of random sampling in small areas. First, we obtain ordinary least squares (OLS) estimates of the element residuals e_{ij} as

$$\hat{e}_{ij\,\mathrm{OLS}} = y_{ij} - \hat{\overline{Y}}_i - \left(\mathbf{X}_{ij} - \hat{\overline{\mathbf{X}}}_i\right)\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

where $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ is the OLS estimate of the regression coefficient $y_{ij} - \hat{\overline{Y}}_i$ regressed on $\mathbf{X}_{ij} - \hat{\overline{\mathbf{X}}}_i$ (no intercept). Only residuals for nonzero values of $\mathbf{X}_{ij} - \hat{\overline{\mathbf{X}}}_i$ are computed. From these residuals we estimate

$$\hat{\sigma}_e^2 = \frac{1}{(n - m - n_0)} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \hat{e}_{ij\mathrm{OLS}},$$

where n_0 is the number of zero-valued **x** residuals. Second, we estimate the OLS residuals (u) from a regression of $y_{ij} \times b_{ij}^{-1}$ on $x_{ij} \times b_{ij}^{-1}$ (no small-area effects), i.e.,

$$\hat{u}_{ij\mathrm{OLS}} = b_{ij}^{-1}\left(y_{ij} - \mathbf{X}'_{ij}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\right),$$

and obtain an estimate of σ_v^2 from

$$\hat{\sigma}_v^2 = \frac{1}{n^*}\left(\sum_{i=1}^{m} \sum_{j=1}^{n_i} \hat{u}_{ij\mathrm{OLS}}^2 - (n-p)\hat{\sigma}_e^2\right),$$

where

$$n^* = \sum_{i=1}^{m} n_i \hat{\overline{\mathbf{X}}}_i \left[1 - n_i \left(\sum_{i=1}^{m} \sum_{j=1}^{n_i} \mathbf{X}_{ij}\mathbf{X}'_{ij}\right)^{-1} \hat{\overline{\mathbf{X}}}_i\right].$$

Summation in these expressions should be limited to small areas with $n_i > 1$. Ghosh and Rao (1994) proposed the following estimator for the MSE:

$$\widehat{\mathrm{MSE}}\left(\overline{Y}_{i\mathrm{EBLUP}}\right) = \left(1 - \frac{n_i}{N_i}\right)^2 \left[\hat{g}_{1i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + \hat{g}_{2i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + 2\hat{g}_{3i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)\right],$$

where

$$\hat{g}_{1i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) = \hat{\gamma}_i\left(\frac{\hat{\sigma}_e^2}{n_i}\right) + \left(1 - \frac{n_i}{N_i}\right)^2 \frac{(N_i - n_i)}{N_i^2},$$

$$\hat{g}_{2i}\left(\hat{\sigma}_{\mathrm{v}}^2,\hat{\sigma}_{\mathrm{e}}^2\right)=\hat{\sigma}_{\mathrm{v}}^2\left[\mathbf{X}_i^{\mathrm{C}}-\hat{\gamma}_i\hat{\bar{\mathbf{X}}}_i\right]\hat{\Lambda}^{-1}\left[X_i^{C}-\hat{\gamma}_i\hat{\bar{\mathbf{X}}}_i\right]',$$

and

$$\hat{g}_{3i}\left(\hat{\sigma}_{\mathrm{v}}^2,\hat{\sigma}_{\mathrm{e}}^2\right)=\frac{1}{n_i^2\times\hat{\gamma}_i^3}\times\left(\frac{2\hat{\sigma}_{\mathrm{e}}^8}{n-m-n_0-2}+\frac{2\hat{\sigma}_{\mathrm{v}}^8}{n^{\star}-2}\right)+\frac{4\left(m-1\right)}{n^{\star}}\hat{\sigma}_{\mathrm{e}}^6\hat{\sigma}_{\mathrm{v}}^2,$$

where $\mathbf{X}_i^{\mathrm{C}}$ is a matrix of nonsampled $\mathbf{x}_{ij}$ values in the ith small area and

$$\hat{\Lambda}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathbf{X}'_{ij}\mathbf{X}_{ij}-\hat{\gamma}_i\times n_i\hat{\bar{\mathbf{X}}}'_i\hat{\bar{\mathbf{X}}}_i,$$

Flores and Martínez (2000) entertained the unit level mixed linear model for the estimation of crop areas under irrigation in 53 small areas in the Duero river basin in northwestern Spain. Auxiliary variables were estimates obtained from remotely sensed images and the element of sampling was a 500-m × 500-m ground area (25 ha). A random sample of 158 elements was taken (0–17 per small area). The use of the auxiliary information resulted in a reduction of the MSE of small-area estimates by 30–70%. Kangas (1996) used the mixed element level model for estimating the timber volumes in eight Finnish municipalities and found it efficient (as opposed to direct or synthetic estimation) even in the absence of auxiliary information. Wang and Fuller (2003) recently suggested some improvements to the MSE estimation procedures of mixed linear models; the improvement makes the MSE more robust when the among-area variation is strong. Interested readers are referred to their text for details.

3.8.5 Small-Area Estimation by Block Kriging

The spatial distance between two population elements can be an indicator of the expected similarity of their attribute values. Within a forest stand, for example, one would expect that the basal area in a 100-m^2 unit, on average, would be more similar to the basal area in units that are spatially close than to the basal area in more distant units. This phenomenon of distance-dependent similarity, if manifest, can be exploited in certain small-area estimation problems. Samples taken in the vicinity of the small area can be used efficiently to predict, via a spatial model (Cressie and Chan 1989), the average attribute value in the small area. Simple kriging is a basic form of spatial prediction for a location with unknown attribute values (Goovaerts 1997). A simple kriging prediction is a linear combination of known attribute values observed in locations within a neighborhood of the location for which we seek a prediction The weight given to an observation depends on the strength of the expected covariance between the observed value and the value to be predicted.

Small-area prediction by block kriging illustrates one of the simplest yet most powerful spatial models for small-area estimation problems. As before, we have a population sample of size n with attribute y and we wish to estimate, say, the mean $\bar{Y}_i$ for small area i. We have n_i sample records available for the small area $i\ (n_i \geq 0)$, which we consider as a "block" in the context of kriging. Let us assume that there is no significant spatial trend in the observed y values in the small area and its vicinity but y values in locations separated by less than a relatively short distance of a few hundred meters do tend to be significantly and positively correlated with each other. Furthermore, we also have a function $C(y_k, y_l)$ that predicts, without bias, the expected covariance between two y values observed in locations k and l. Normally the covariance is governed by the distance between locations k and l. Issues surrounding the selection, estimation, and validation of the spatial covariance function (variogram) are beyond the scope of this text. Suffice to say that there are many complex statistical issues one must consider before accepting a spatial model, as the risk of inadvertently introducing a serious bias is nontrivial (Cressie 1991; Chilès and Delfiner 1999; Atkinson and Lewis 2000; Diblasi and Bowman 2001; Zhao and Wall 2004).

In our chosen variant of block kriging the n_i sample records from the small area are used only for a direct estimate $\hat{\bar{Y}}_i$, which is then combined with a block kriging prediction $\hat{\bar{Y}}_{i\,\text{bkrig}}$ in the form of a composite estimator. Attribute values, both sampled and nonsampled, for the small area i (SA_i) are denoted by y_i. The first step towards obtaining $\hat{\bar{Y}}_{i\,\text{bkrig}}$ is to choose a number of sample records $Y_k, k = 1, \ldots, N_i^{\text{OK}}$ taken outside the small area but close enough to the small area to justify the expectation that their attribute values would be significantly correlated with attribute values inside the small area. The covariance function can guide the cutoff distance for N_i^{OK} since $C(y_k, y_i) \times \text{var}(y)^{-1}$ can be viewed as a crude predictor of the expected correlation of y values. Since block kriging computations increase with the square of N_i^{OK} there are good reasons to keep the number as low as possible but high enough to take advantage of spatial correlation. In practice, sample locations with an expected correlation below 0.2 can be excluded with only a minimal impact on the predictions and their estimated variance. The set of outside sample values included for block kriging prediction is $\Im_i^{\text{OK}}$. From the selected outside sample points we obtain the block kriging prediction for the small area as

$$\hat{\bar{Y}}_{i\text{bkrig}} = \sum_{k=1}^{N_i^{\text{OK}}} \hat{\lambda}_k \times Y_k, Y_k \in \Im_i^{\text{OK}},$$

where $\hat{\lambda}_k$ is the estimated block kriging weight assigned to Yk in $\Im_i^{\text{OK}}$. Estimated block kriging weights are obtained as solutions to the following system of block kriging equations (Goovaerts 1997):

$$\sum_{k=1}^{N_i^{\mathrm{OK}}} \lambda_k \times C(Y_k, y_l) = \overline{C}(y_l, y_i),\ y_i \in SA_i,\ \{Y_k, y_l\} \in \Im_i^{\mathrm{OK}},$$
$$\sum_{k=1}^{N_i^{\mathrm{OK}}} \lambda_k = 1,$$
$$l = 1, \ldots, N_i^{\mathrm{OK}},$$

where $\overline{C}(y_h, y_l)$ is the average expected covariance between Y_k from $\Im_i^{OK}$ and an element attribute value y_l in SA_i. We can obtain a good approximation of $\overline{C}(Y_k, Y_l)$ by computing the expected covariance between elements in $\Im_i^{OK}$ and a series of elements y_l, $l = 1, ..., M_{\mathrm{SA}_i}$ distributed evenly over SA_i and then take the average of these expected covariances (Goovaerts 1997); hence,

$$\overline{C}(y_l, y_i) \simeq \frac{1}{M_{\mathrm{SA}_i}} \sum_{i=1}^{M_{\mathrm{SA}i}} C(y_l, y_i),\ y_l \in \Im_i^{\mathrm{OK}},\ y_i \in \mathrm{SA}_i,$$

The number of locations M_{SA_i} to be included in the average depends on the rate of convergence. At one point increasing the number further will have only a minor impact on the average. Around 16 is probably a reasonable choice. An estimator of the variance of $\hat{\overline{Y}}_{i\,\mathrm{bkrig}}$ is

$$\hat{\mathrm{var}}\left(\hat{\overline{Y}}_{i\,\mathrm{bkrig}}\right) = \overline{\overline{C}}(Y_k, y_j) - \sum_{k=1}^{N_i^{\mathrm{OK}}} \hat{\lambda}_k \times \overline{C}(Y_k, y_l),\ \{y_i, y_k\} \in \mathrm{SA}_i,$$

where $\overline{\overline{C}}(y_i, y_j)$ is the expected block-to-block covariance which we approximate by the average covariance between elements in SA_i, i.e.,

$$\overline{\overline{C}}(Y_k, y_l) - \sum_{i=1}^{M_{\mathrm{SA}i}} \sum_{l=1}^{M_{\mathrm{SA}i}} C(Y_k, y_l),$$

The block kriging predictor $\hat{\overline{Y}}_{i\,\mathrm{bkrig}}$ and the direct estimate $\hat{\overline{Y}}_i$ can now be combined to a composite estimator as outlined in Sect. 3.8.3. To compute the MSE of the composite estimator we need an estimate of the expected covariance between $\hat{\overline{Y}}_{i\,\mathrm{bkrig}}$ and $\hat{\overline{Y}}_i$. They cannot be assumed to be independent since $\hat{\overline{Y}}_{i\,\mathrm{bkrig}}$ implicitly generates M_{SA_i} pseudo-observations for SA_i. The expected covariance is approximated by the average expected covariance between the n_i sample locations in SA_i and the N_i^{OK} sample locations in $\Im_i^{\mathrm{OK}}$. Alternatively, a single prediction $\hat{\overline{Y}}_{i\,\mathrm{bkrig}}$ could have been obtained by including the n_i sample points in SA_i in the set $\Im_i^{\mathrm{OK}}$ with no other change to the previously outlined procedure. We chose a composite estimator as it is more transparent and computationally easier to optimize. The choice will have only a minor impact on the results.

When y values display a spatial trend the block kriging procedure has to be expanded to include the prediction of local trend values. While the extension is technically straightforward the presence of a trend nevertheless complicates matters. First, the trend has to be estimated precisely to avoid introduction of

potentially serious bias. A precise estimation is often not possible from typical survey data. Also, only few surveys will have enough data to support a thorough model selection and validation process and a high enough sampling density to ensure that there are a sufficient number of suitable predictors available in the area around SA_i (Lappi 2001). Still, there are situations in forest inventory where spatial-model-based small-area predictions are attractive (Mandallaz 1993, 2000). They are identified by the presence of a significant distance- or location-dependent correlation, viz., covariance, between sampled attribute values.

3.8.6 Empirical Bayesian Methods for Small-Area Estimation

The Bayesian framework uses the posterior distribution for inference (Ghosh and Meeden 1997; Gaudard et al. 1999; Congdon 2001). The posterior distribution is the product of a likelihood and a prior distribution, and as such it is entirely model-based. A likelihood function f with parameters $\boldsymbol{\theta}_i$ can be postulated for the data sampled in a small area i (SA_i) and then combined with prior expectations with regard to the probability distribution function of the parameters $\boldsymbol{\theta}_i$ in order to obtain the posterior distribution p of $\boldsymbol{\theta}_i$. When prior distributions are estimated from samples taken outside SA_i p is said to be the empirical Bayesian (EB) posterior for SA_i (Singh et al. 1998; Pfeffermann 2002). The EB approach offers a very flexible and rich framework for small-area estimation. In applications with a Gaussian-likelihood function and a conjugate prior (a conjugate prior produces a posterior distribution of the same type as implied by the likelihood), the posterior estimates will be similar to the composite estimator (Congdon 2001).

An example with a continuous real-valued positive attribute y and one with counts of a categorical attribute value illustrate the flexibility and power of the EB approach. In our first EB example, suppose we have from SA_i four (n_i=4) sample values $Y_i = \{Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}\}$={156, 220, 181, 185} with a mean $\hat{\bar{Y}}_i$ of 185.5 and a variance $\hat{\mathrm{var}}\left(\hat{\bar{Y}}_i\right)$ of 173.4. A larger sample of size $n - n_i = 100 - 4 = 96$ from outside of SA_i produced $\hat{\bar{Y}}_{\supset SA_i} = 200$ and $\hat{\mathrm{var}}\left(\hat{\bar{Y}}_{\supset SA_i}\right) = 4.59$. The small-area likelihood f is a Gaussian with $\boldsymbol{\theta}_i = \{\theta_{i1}, \theta_{i2}\} = \left\{\bar{Y}_i, \mathrm{var}\left(\bar{Y}_i\right)\right\}$ and we seek to estimate $\hat{\boldsymbol{\theta}}_i^{\mathrm{EB}}$ the EB posterior of $\boldsymbol{\theta}_i$. We assume a Gaussian prior with parameters $\hat{\omega} = \left\{\hat{\bar{Y}}_{\supset SA_i}, \hat{\mathrm{var}}\left(\hat{\bar{Y}}_{\supset SA_i}\right)\right\}$ for the mean and a Γ (gamma)distribution prior with parameters $\hat{\lambda} = \{96.00, 0.048\}$ for the variance. The Γ distribution prior was chosen so that its expected value would be $\hat{\mathrm{var}}\left(\hat{\bar{Y}}_{\supset SA_i}\right)$ and its variance $2 \times \hat{\mathrm{var}}^2\left(\hat{\bar{Y}}_{\supset SA_i}\right) \Big/ (n - n_i - 2)$, which is the variance of a variance when y is normally distributed (Snedecor and Cochran 1971). From these preliminaries we obtain the posterior p of $\boldsymbol{\theta}_i$ as

$$p\left(\boldsymbol{\theta}_i \middle| \hat{\bar{Y}}_i\right) \propto f\left(y_{ij} \middle| \boldsymbol{\theta}_i\right) \times \varphi\left(\theta_{i1} \middle| \hat{\omega}\right) \times \Gamma\left(\theta_{i2} \middle| \hat{\lambda}\right).$$

The maximum posterior log-likelihood was −10.87 with $\boldsymbol{\theta}_i^{\mathrm{EB}} = \left\{\hat{\bar{Y}}_{i\,\mathrm{EB}}, \hat{\mathrm{var}}\left(\hat{\bar{Y}}_{i\,\mathrm{EB}}\right)\right\}$ = {192.8, 9.7}. In this example we gave the sample from outside SA_i maximum weight, in that our priors were tailored to the large-sample results. The EB posterior mean is almost a perfect average of $\hat{\bar{Y}}_{\neg \mathrm{SA}_i}$ and $\hat{\bar{Y}}_i$ (an estimated weight of 0.49 is given to the direct estimate) and the posterior variance is almost 18 times smaller than the variance of the direct estimate, but also about 18 times as large as the variance of the $\hat{\bar{Y}}_{\neg \mathrm{SA}_i}$. We are of course free to change the priors if only a part of the large sample is deemed representative as a prior for SA_i or we have other information that warrants a change. In any case, the choice of an informative "prior" must be decided carefully and should be justified explicitly in the same fashion as one would justify a model choice. The EB framework is extended easily to deal with regression and ratio estimators and multivariate attribute values (Ghosh and Meeden 1997; Green and Valentine 1998; Elliot and Little 2000; Denison et al. 2002; O'Brien and Dunson 2004).

Our second example show the flexibility of the EB appraoch to handle binary data. We have done a survey of a beetle infestation. In each plot n_t=16 trees are examined for the presence (y=1) or absence (y=0) of a certain beetle species. We have a total of n=47 plots, of which n_i=7 are inside SA_i. We wish to estimate the proportion of trees infested with the beetle in the small area (P_i) and a variance of this estimate. At the plot level, the likelihood of observing, say, $n_{\mathrm{b}j}$ beetle-infested trees in the jth plot is

$$\Pr\left(n_{\mathrm{b}j} \middle| n_t, P_i\right) = \binom{16}{n_{\mathrm{b}j}} \times P_i^{n_{\mathrm{b}j}} \left(1 - P_i\right)^{n_t - n_{\mathrm{b}ji}}$$

as per the binomial distribution. The results from the 40 "outside" plots are used to form prior expectations of the proportion P_i. The survey produced the following estimates:

$$\hat{P}_{\supset \mathrm{SA}_i} = 0.172, \hat{\mathrm{var}}\left(\hat{P}_{\supset \mathrm{SA}_i}\right) = 5.196 \times 10^{-4}$$

for the outside area and $\hat{P}_i = 0.223$, $\hat{\mathrm{var}}\left(\hat{P}_i\right) = 2.949 \times 10^{-3}$ for SA_i, where 25 of the 112 sample trees were infested with the beetle. We assume conveniently a beta distribution as a prior for the parameter P_i in the small-area data likelihood. The beta distribution has two parameters, α and β and a mean of $\bar{P} = \alpha \times (\alpha + \beta)^{-1}$ and a variance of $\bar{P}\,(1-\bar{P}) \times (1 + \alpha + \beta)^{-1}$. From the 40 outside plots and by maximum-likelihood methods, we obtained $\hat{\alpha} = 1.005$ and $\hat{\beta} = 4.843$. The convenience in the choice of the prior is that the posterior distribution of the small-sample estimate of P is also a beta distribution (Congdon 2001) but with parameters $\left\{\hat{\alpha} + \sum_{i=1}^{n_i} n_{\mathrm{b}i},\ \hat{\beta} + n_i \times n_t + 1\right\} = \{26.00, 117.84\}$, from which we obtain $\hat{P}_{i\mathrm{EB}} = 0.221$ and $\hat{\mathrm{var}}\left(\hat{P}_{i\mathrm{EB}}\right) = 1.44 \times 10^{-3}$. While the posterior

mean is close to the direct sample mean, the EB estimate of the posterior variance is only half of the variance of the direct estimate.

EB extensions to multinomial data are straightforward: instead of a beta distribution as a prior one conveniently chooses instead the Dirichlet distribution (Santner and Duffy 1989; Green and Clutter 2000), which generates a Dirichlet posterior with parameters determined as the sum of small-area counts and the prior parameters (pseudo-counts).

3.9 k Nearest-Neighbor Prediction

When an auxiliary attribute(s) is known for all N population elements and a functional relationship can be assumed to exist between them and the attribute of interest, available for $n\,(n \ll N)$ sampled elements only, then the predictive power of the auxiliary attribute values can be exploited in several ways for the purpose of improving the precision of an estimated mean or total of a population or a stratum. This was illustrated in Sect. 3.3.5 for two-phase sampling. While global and strata estimates of totals and means are useful in their own right, the management of natural resources often requires attribute values to be provided for all population elements within specified areas. Essentially a map showing the spatial distribution of attribute values is desired. A naive prediction of local attribute values from a population-level regression model can produce unacceptable local artefacts because the predictions ignore any spatial correlation among the predictors and because application of a single population-level model may produce biased results when applied to spatial subsets of the population (Rao 2003).

A forest can be viewed as a composition of a finite set of distinct compositions of auxiliary attribute values. If, furthermore, we assume that the value of the attribute(s) of interest is fixed for a given distinct composition of the auxiliary value(s) then, if the assumption holds, perfect predictions would be possible when the distinct set of auxiliary compositions matches that of the entire population. The predicted attribute value would naturally be the value recorded for the sample with matching auxiliary values. In practice a perfect match is rarely possible because the sample simply does not exhaust the natural variability in the auxiliary attribute(s). To make our predictions we could relax our requirement of a perfect match and assume that similar auxiliary attribute values means similar values of the desired attribute. The *k*NN method of prediction is based on this relaxed assumption and was first developed for the purpose of replacing within-item missing attribute values (Rubin 1987). In the *k*NN method a prediction is derived from the k sample records that match most closely the auxiliary values of the element we wish to predict.

The *k*NN method of prediction is intuitively appealing and conceptually simple; however, a successful application demands complex and demanding analyses and computations. The numbers of neighbors, the auxiliary traits to include, and the definition of similar values are all nontrivial issues in need of careful analysis (Moeur et al. 1995; Katila and Tomppo 2001; McRoberts et al. 2002).Otherwise, carefully calibrated *k*NN predictions will be biased and the result could be worse than predictions based on global expectations (Holmgren et al. 2000; Franco-Lopez et al. 2001). *k*NN is now used routinely to provide local estimates of national forest inventory attributes from local auxiliary attribute values obtained from remotely sensed images (Gjertsen et al. 1999; Katila et al. 2000; Katila and Tomppo 2001; Tomppo and Halme 2004).

All *k*NN methods require large sample sizes to ensure that similar matches are indeed found. It is difficult to make specific recommendations, the natural variability in attribute values is the decisive factor, but even for rather homogeneous forests of northern climes sample sizes over 200 seem to be required for *k*NN methods to be even modestly successful (Haara et al. 1997; Franco-Lopez et al. 2001; Holmström 2002). As the number of auxiliary attributes increase it becomes increasingly difficult to find a good match, a paradigm known as the curse of dimensionality (Scott 1992). Predictions derived from the single most similar set of auxiliary attribute values are asymptotically unbiased and they will preserve the sample variability in the desired attribute value(s) (Moeur et al. 1995; McRoberts et al. 2002). When more than one similar sample record is used for prediction, then it is a common observation that predictions at the extreme tend to be biased in opposite directions (Moeur et al. 1995; McRoberts et al. 2002).

The local kNN prediction of the attribute y from the auxiliary variables $\mathbf{X}$ for the ith nonsampled population element is

$$\hat{Y}_i = \sum_{j \in NN_k(\mathbf{x}_i)} w_{ij} \times y_j,$$

where y_j is the attribute value of the jth sample, w_j is the weight given to this value, and j is one of k NNs to the ith population element in terms of the auxiliary attribute values, $j \in NN_k(\mathbf{X}_i)$. In the multivariate case a scalar would be replaced by the appropriate vector notation. The weights are chosen to reflect the degree of similarity in the auxiliary attribute values between the ith nonsampled and the jth sampled population element. Weights are usually based on an index (d_{ij}) of similarity, viz., distance, between the auxiliary attribute values of the ith and jth elements. Subject knowledge, prior beliefs, ecology, spatial distance, and statistical consideration guide the choice of weight function. For example, a close match in $\mathbf{X}$ could still receive a low weight if population

elements i and j are located on different soil types, on different aspects, in different vegetation zones, or are separated by a large spatial distance. Prior knowledge can be used to guide the search towards locations with the highest chance of a suitable match (van Lieshout and Baddeley 2002). It is customary to choose weights such that

$$w_{ij} = \frac{1}{d_{ij}^t} \times \left(\sum_{j \in NN_k(\mathbf{x}_i)} \frac{1}{d_{ij}^t} \right)^{-1}, t = 0,1,2,\ldots.$$

t=0 implies equal weighting of the k neighbors, whereas t=1 weights with the inverse of distance and higher t values ensure a more rapid decline in the weight is given to the less similar values in the group of k most similar neighbors. The choice of t is intimately connected to the number k of most similar neighbors. It makes little sense to have a high k and a high t as most neighbors would then contribute little towards a prediction. Conversely, a lower k would argue for a lower t. A t value of 1 seems to be the most popular choice.

The similarity index d_{ij} should reflect the impact that discrepancies in the auxiliary variable(s) have on a local prediction. An ideal index is linear in the square of the absolute prediction errors (Barbieri and Berger 2004). The index is inevitably a function of the auxiliary attributes included as predictors, their scale, and predictive power. Finding an optimal index or distance metric is the crux of the *k*NN method and is often a very time consuming step. A generic index takes the form

$$d_{ij} = \left(\mathbf{x}_i - \mathbf{x}_j\right) \Omega_{\mathbf{x}'\mathbf{x}}^{-1} \left(\mathbf{x}_i - \mathbf{x}_j\right)',$$

where $\mathbf{x}$ is a $p \times 1$ vector of auxiliary attribute values, $\Omega_{\mathbf{x}'\mathbf{x}}$ is a weight matrix, and $\mathbf{x}'$ is the transpose of $\mathbf{x}$. In the case of $\Omega_{\mathbf{x}'\mathbf{x}} = \mathrm{I}$, the identity matrix, the similarity index is equal to the Euclidian distance in the feature space of $\mathbf{x}$. A Euclidian distance weighting disregards the predictive power of individual auxiliary attributes and distances are strongly influenced by scale differences in the auxiliaries. The choice of $\Omega_{\mathbf{x}'\mathbf{x}} = D\left(\sigma_j^2\right)$, where $D\left(\sigma_j^2\right)$ is a diagonal matrix of the variances of the auxiliary variables, removes the scale effect on the distance measure but does not reflect a possible correlation among the predictors. Disregarding the correlation can lead to biased predictions. Choosing $\Omega_{\mathbf{x}'\mathbf{x}} = \Sigma_{\mathbf{x}'\mathbf{x}}$, where $\Sigma_{\mathbf{x}'\mathbf{x}}$ is the variance covariance matrix of the auxiliary attributes, leads to a similarity index based on Mahalanobis distances (Rencher 2002) and removes both scale effects and correlation between the auxiliary attributes, but their predictive powers is not taken into account. These choices of the weight matrix result in nonparametric *k*NN predictions. The predictive power of the auxiliary attributes can be

incorporated by assuming that predictions of y are linear in $\mathbf{x}$, i.e., $\hat{y} = \hat{\boldsymbol{\beta}}\mathrm{x}'$, or are linear in a set of mutually independent (orthogonal) variables $\mathbf{z}$ obtained by premultiplying $\mathbf{x}$ by the Cholesky decomposition of $\Sigma_{\mathrm{x}'\mathrm{x}}$ (Rencher 2002). In the former case we get

$$d_{ij} = \left(\mathbf{x}_i - \mathbf{x}_j\right)\hat{\boldsymbol{\beta}}\Omega_{x'x}^{-1}\hat{\boldsymbol{\beta}}'\left(\mathbf{x}_i - \mathbf{x}_j\right)',$$

where $\hat{\boldsymbol{\beta}}$ is either an ordinary or a generalized least-squares estimate of regression coefficients, and in the latter case we get

$$d_{ij} = \left(\mathbf{x}_i - \mathbf{x}_j\right)\Gamma\Lambda^2\Gamma'\left(\mathbf{x}_i - \mathbf{x}_j\right)',$$

where Γ is a matrix of canonical correlation coefficients and $\Lambda_{z'z}$ a diagonal matrix of canonical correlation coefficients (Rencher 2002). The two distance measures are identical if all p of the transformed variable $\mathbf{z}$ are used as predictors. If only a subset q ($p>q$) with a significant correlation to y (or $\mathbf{y}$) is used then the two will differ. Further details on the canonical approach can be found in Moeur et al. (1995).

The expected error of a kNN-predicted value of y is usually estimated by some leave-one-out cross-validation procedure (Franco-Lopez et al. 2001; McRoberts et al. 2002; Rao 2003; Efron 2004). The procedure is relatively simple but time-consuming. The one-by-one procedure makes a kNN prediction $\hat{y}_{k\mathrm{NN}}$ for one of the n sampled elements by withholding this observations from the calculations of similarity indices, weights, and ranking of indices. The mean of the errors made in these n predictions is the cross-validation estimate of error:

$$\mathrm{RMSE}_{\mathrm{CV}}\left(\hat{Y}_{NN_k}\right) = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_{NN_k}^{(i)}\right)^2}{n}},$$

where $\hat{Y}_{NN_k}^{(i)}$ is a kNN prediction of the ith sample value derived independently from Y_i. Bootstrapping offers an alternative method for estimating this error. Instead of the delete one-at-a-time procedure of cross-validation, a sample of size n is sampled with replacement from the original sample and a kNN prediction rule is obtained from the bootstrap sample and is applied to the original sample. By repeating this process a large number of times (more than 500) one obtains a distribution of prediction errors from which statistics such as the mean, mode, and quantiles are easily obtained. Estimates of the prediction variance are only approximations, possibly biased since predictions are based on order statistics with a nonsmooth distribution function (Chen and Shao 2001).

Individual kNN prediction errors can be assumed to depend on the index of similarity values d_{ij} for the kNNs used for a prediction; hence, a regression model with the square of the prediction errors obtained during the cross-validation process as the dependent variable, and the k d_{ij} index values as the

predictors could be used to estimate the *k*NN error of individual predictions (Moeur et al. 1995).

3.10 Resampling for Nonlinear Inventory Statistics

Users of forest inventory information are often interested in estimates that go beyond the mean, the total, and associated estimates of their sampling variance. Estimates of, for example, population percentiles (e.g., the median and the lower and upper 2.5 percentiles), ratios of estimates involving two or more inventory attributes (e.g., percentage change during a given period of time or the proportion of area in plantations), number of species in a population, or simply a model-based transformation of one or several inventory estimates into another attribute of interest (e.g., composite estimators, transformation of volume to biomass or carbon content, small area estimates, estimates of nonsampling errors) are demanded on a routine basis from the analyst. Let $\hat{T}$ be such an estimate obtained from one or several inventory estimates $\hat{\mathrm{Z}}$ via some function g as in $\hat{T} = g\left(\hat{\mathrm{Z}}\right)$, where $\hat{\mathrm{Z}} = \left\{\hat{Y}_1, \ldots, \hat{Y}_k, \left\langle \hat{\mathrm{cov}}\left(\hat{Y}_i, \hat{Y}_j\right)(i,j) = 1, \ldots, k \right\rangle\right\}$. If g is linear in the inventory estimates (as in a weighted average with fixed and known weights), the variance of $\hat{T}$ is estimated via a first-order Taylor series linearization-substitution method (Rao 1988):

$$\hat{\mathrm{var}}\left(\hat{T}\right) = \hat{g}'\left(\mathrm{Z}\right)^{T} \hat{\boldsymbol{\Omega}}\left(\mathrm{Z}\right) \hat{g}'\left(\mathrm{Z}\right)$$

where $g'\left(\mathrm{Z}\right)$ is the vector of derivatives with respect to the inventory attributes and $\hat{g}'\left(\mathrm{Z}\right)$ is $g'\left(\mathrm{Z}\right)$ evaluated at $\hat{\mathrm{Z}}$, $\hat{\boldsymbol{\Omega}}\left(\mathrm{Z}\right)$ is the estimate of the variance covariance matrix of **Z**, and superscript T denotes the transpose of a vector or matrix.

For g linear in all the parameters **Z** the estimate $\hat{T}$ and the estimate of the variance of T will exhibit properties that are a linear function (g) of the elements of **Z**. If $\hat{\boldsymbol{Z}}$ is design-unbiased so is $\hat{T}$, and if the variance estimates for **Z** are all design-unbiased and consistent so is the estimated variance of T. However, when g is nonlinear, or possibly nonsmooth (derivatives do not exist everywhere, as in a discrete distribution or when g embodies a series of hierarchical functions or the output of g is constrained), the statistical properties of $\hat{T}$ are no-longer predictable from g and the properties of $\hat{\boldsymbol{Z}}$. $\hat{T}$ may be biased and the Taylor series method may produce a poor approximation to the variance of T since higher moments of the sampling distribution of T do not vanish (as they do in a normal distribution).

When g is nonlinear or nonsmooth the analyst may chose to adopt a resampling scheme as an alternative to the Taylor series method. Research has shown

that estimates of $g\left(\hat{Z}\right)$ and $\hat{\text{var}}\left[g\left(\hat{Z}\right)\right]$ with g nonlinear or nonsmooth derived from a correct application of the bootstrap resampling technique are at least as good as those based on the Taylor series method; often they are better (Shao 1996, 2003; Shao and Chen 1998; Hall et al. 2001; Van Hees 2002; Lahiri 2003; Shen et al. 2004; Zhu and Morgan 2004). At times the function g is so complex as to preclude an analytical estimator that the bootstrap or another resampling alternative (e.g., jackknife, balanced replicate resampling, or Pòlya-urn) offers the only practical option (Shao 1996; Meeden 1999). Estimation of the MSE of composite estimators as exemplified by small-area estimation problems (see also Sect. 3.8) becomes almost straightforward with bootstrap resampling. Bootstrap resampling is also attractive for estimation problems when missing data are imputed at random (hot-deck) or model-based (Saho and Sitter 1996; van Deusen 1997; Shao and Steel 1999; McRoberts 2001; Lahiri 2003; Shao 2003). Although computer-intensive, the bootstrap computations are simple.

3.10.1 The Bootstrap

Efron (Efron and Tibshirani 1993) introduced the bootstrap resampling method for the study of the properties of no-linear and nonsmooth statistics. The bootstrap simulates the estimated sampling distribution of a statistic estimating a population attribute by generating a large number (B) of bootstrap estimates $\hat{T}_1^*, \ldots, \hat{T}_B^*$ of T from which a mean, a variance, and an approximation to the distribution function $\hat{\Pr}\left(\hat{T}^* \leq T\right)$ are obtained by standard techniques.

In the simplest (naive) implementation of bootstrap resampling a single bootstrap estimate $\hat{T}_l^*$, $l = 1, \ldots, B$, is obtained by a SRS *with replacement* from n observed values of Y_i, $i = 1, \ldots, n$. The resampling yields a bootstrap sample Y_j^*, $j = 1, \ldots, n$ from which $\hat{T}_l^*$ is estimated. The naive implementation requires that the observed Y_i are identically and independently distributed (iid), which is only possible if the data are collected by SRS. Sample selection with unequal probabilities, however, invariably introduces a complex correlation structure which makes the development of a theoretical valid bootstrap method challenging (Lahiri 2003). For sample surveys, bootstrapping methods have been validated under randomization theory.

In forest inventories sampling is commonly from a finite population and without replacement to avoid sampling the same element (unit) more than once. Even point-sampling locations are usually chosen amongst a finite set of possible locations. Consequently variance estimators include a correction factor for the sampling fraction (f) in a finite population and the variance-effective sample size under sampling without replacement is n–1 not n as in sampling with replacement (Thompson 1992). These differences, if not carefully

identified and accounted for in the bootstrap resampling procedure, can lead to a problem of bias in bootstrap estimates of variance and percentiles (Lahiri 2003; Shao 2003). Schreuder and Williams (2000) found conventional 95% confidence intervals for the mean under SRS and sample sizes of 20, 40, and 60 to be slightly superior in terms of actual coverage of the true mean than corresponding naive bootstrap confidence intervals.

A large number of modified bootstrap procedures have been proposed to account for the sampling procedure and finite-population corrections (Shao 1996). The bootstrap can adapt to any sampling design with the provision that resampling is done at the unit level (h) at which the *iid* assumption is still valid given that the unit was sampled. In stratified multi-stage cluster sampling, for example, bootstrap resampling would occur at the level of clusters within strata. Attribute values in a cluster can not be assumed *iid* conditional on inclusion of the cluster in the sample.

Common features of modified bootstrap procedures are (1) the resample size m_h is less than the size of the available sample n_h at unit level h of resampling, (2) a scaling of $\mathbf{y}$, and (3) resampling without replacement from a synthetic "complete" population with the sample records multiplied from n_h to N_h. Only the flexible rescaling bootstrap proposed by Rao and Wu (1988) will be detailed here for the case where the actual sample units were selected by SRS or unequal inclusion probabilities.

Under the SRS scenario at unit level h one obtains a bootstrap sample Y^*_{hj}, $j=1,\ldots,m_h < n_h$, by SRS with replacement which is then rescaled according to

$$\tilde{Y}^*_{hi} = \overline{Y}_h + \sqrt{\frac{m_h(1-f_h)}{n_h-1}}\left(Y^*_{hj} - \overline{Y}_h\right),\ j = 1,...,m_h,$$

where $\overline{Y}_h$ is the mean of the actual sample at unit level h and f_h is the actual sample fraction in unit level h (by count or area). This process is repeated for all unit levels h (h=1,....,H). Then, one obtains a design-based estimate $\tilde{T}^*_l$ of T as if $\tilde{Y}^*_{hi}$, $j=1,\ldots m_h$, $h=1,\ldots,H$, was an actual sample. B replications of this process produce the rescaled bootstrap estimate of the sampling distribution of T. Under an unequal probability sampling design the jth element in unit level h is given a weight w_{hj} in order to expanded it to an (unbiased) estimate of the population total. The bootstrap resampling is done as under SRS but instead of rescaling Y^*_{hj} one rescales the sampling weights

$$w^*_{hj} = w_{hj}\left(1 - \sqrt{m_h \times (n_h-1)^{-1}} + \sqrt{m_h \times (n_h-1)^{-1}} \times \frac{n_h}{m_h} \times r^*_{hj}\right),$$

where r^*_{hj} is the number of times Y_{hi} is included in the bootstrap resample. After completing the bootstrap resampling across all H unit levels the desired sample estimate is obtained from Y^*_{hj} using weights w^*_{hj} in place of w_h (Rao et al. 1992).

Modified bootstrap procedures extend naturally to estimation with missing data replaced by random or model-based imputations (Rubin 1987). The bootstrap sample is taken from the complete nominal sample. Missing sample records are simply added as empty records with a label that identifies them as missing. After each round of modified bootstrap resampling the missing values are imputed from the drawn bootstrap sample (only) by applying the exact same protocol as would be used for the actual sample. If one did the imputations before the bootstrapping the variance estimates would be downward biased (Shao 1996).

A major advantage of the bootstrap of multivariate data is the effortless provision of measures of multivariate associations within and between sampling units since these are estimated by standard procedures from the replicated bootstrap samples. These associations are almost always needed for the estimation of variances of complex survey estimators. Needless to say, conventional methods for obtaining estimates of these quantities can be exceedingly difficult.

3.10.2 The Jackknife Resampling

The jackknife is a delete n units at a time resampling technique (Efron 1982) used to obtain first-order approximations to estimates of bias, and sampling variance. When the function g is linear, the jackknife estimates will be equivalent to the design-based estimates. When g is nonlinear or nonsmooth, a jackknife variance estimator may be inconsistent (Shao 1996). As for the bootstrap, the rationale for using a jackknife resampling procedure for estimation is not statistical but rather convenience when design-based or model-based estimators are exceedingly complex or nonexistent.

When the data are a simple random sample of $y_i, i = 1, ..., n$, the ith leave-one-out jackknife sample is

$$y_{(i)} = \{Y_1, Y_2, ..., Y_{i-1}, Y_{i+1}, ...Y_n\},$$

from which the ith jackknife replication of $\hat{T}_{(i)} = g\left(Y_{(i)}\right)$ is obtained as for the actual sample. After obtaining possible distinct all n jackknife replication estimates of T the jackknife estimate of bias in the estimate $\hat{T}$ obtained from the actual sample is

$$\text{bias}_{\text{jack}} = (n-1)\left(\hat{T}_{(\cdot)} - \hat{T}\right) \text{ with } \hat{T}_{(\cdot)} = n^{-1}\sum_i T_{(i)}$$

and the jackknife estimate of the standard error is

$$\hat{\text{var}}_{\text{jack}}\left(\hat{T}\right) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{T}_i - \hat{T}_{(\cdot)}\right).$$

In finite-population sampling a correction for the sample fraction must be reflected in the jackknife variance estimator. In multistage inventory one has to decide what the unit to delete is. As for the bootstrap, one should delete the highest unit level unit for which the iid assumption is valid given the unit is included in the actual sample. Also, under unequal probability sampling the effect of deleting one unit on the expanded (weighted) sample observations must be addressed. We illustrate the jackknife procedure for unequal probability stratified cluster sampling. Notation is as per the bootstrap example given before. After deleting the j th unit in the h th unit level we obtain the $h\,j$ th jackknife replication of $\boldsymbol{Y}$ from

$$\hat{\mathrm{Y}}_{(h'j')} = \sum_{h \neq h'} \sum_{j \neq j'} w_{hj} \times Y_{hj} + \frac{n_{h'}}{n_{h'} - 1} \sum_{j \neq j'} w_{h'j} \times Y_{h'j},$$

from which we obtain $\hat{T}_{(h'j')}$ as before. We repeat this delete-one process across all units j' in a unit at level and across all unit levels h' (h'=1,...,H) and obtain

$$\hat{T}_{(h')} = n_{h'}^{-1} \sum_{j'=1}^{n_{h'}} \hat{T}_{(h'j')}$$

and finally the jackknife variance estimator for $\hat{T}$ from

$$\hat{\mathrm{var}}_{\mathrm{jack}}(\hat{T}) = \sum_{h=1}^{H} \frac{(1 - f_h)(n_{h'} - 1)}{n_{h'}} \sum_{j'=1}^{n_{h'}} \left(\hat{T}_{(h'j')} - \hat{T}_{(h')} \right)^2.$$

3.10.3 The Pòlya-Urn Resampling Scheme

The ease of implementation of the flexible and simple Pòlya-urn resampling scheme and the fact that Pòlya-urn estimators of means, totals, and variances are design-consistent and asymptotically equivalent to design-based estimators (Ghosh and Meeden 1997) makes Pòlya-urn resampling an attractive alternative to the bootstrap. Pòlya-urn resampling generates a posterior distribution of the statistic of interest. It is a predictive joint distribution for the unobserved or unseen units in the population conditional on the seen sample values, similar to the Bayesian bootstrap of Rubin (1981). Problems associated with the sampling process, unequal inclusion probabilities, and finite populations are not encountered in the Pòlya-urn resampling scheme.

The basic Pòlya-urn resampling scheme is very simple. Let us assume that we have n sample observations Y_i, $i = 1, \ldots, n$, from a finite population of size N. To implement the Pòlya-urn sampling we place the n sample records in a virtual urn. We draw one sample record at random from the urn, and return the record and one additional copy of this record to the urn. There are now n+1 sample records in the urn. We repeat this drawing scheme a total of N–n times. After the last draw, the urn contains N sample records, which we interpret as

one posterior prediction of a population census. From the N sample records, we compute the attribute of interest by standard techniques, just as in a bootstrap. A large number K ($K > 400$) of posterior predictions are obtained in this manner. As for the bootstrap, the number K is determined by the among-replicate variability of the posterior predictions.

The Pòlya-urn resampling scheme adapts well to more complex designs. In one-stage cluster sampling the resampling scheme is unchanged; a sample record is the data from a cluster (Magnussen et al. 2004). Under a stratified random sampling design, a Pòlya-urn resampling scheme such as the one just described is implemented for each strata (Magnussen and Köhl 2002). In multistage cluster sampling (Meeden 1999) the resampling is done in a nested sequence. For example, we have under SRS sampled n first-stage clusters out of a total of N, and we have sampled m out of M ultimate units within each cluster. We would then do an urn resampling of the n first-stage units until the urn contained N such units. Then, we would take each first-stage unit in the urn and conduct a second-stage round of urn resampling until the chosen unit contained M ultimate units. After one completion of the nested urn resampling scheme we have one posterior prediction of a population census and we can proceed to compute the statistics of interest. And we continue until additional replications of the nested resampling only produce minimal gains.